REGULATION OF VIRULENCE IN THE PLANT PATHOGEN *PSEUDOMONAS SYRINGAE* PV. *TOMATO* DC3000

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Hanh Ngoc Lam

August 2014

REGULATION OF VIRULENCE IN THE PLANT PATHOGEN *PSEUDOMONAS SYRINGAE* PV. *TOMATO* DC3000

Hanh Ngoc Lam, Ph. D.

Cornell University 2014

The type III secretion system (T3SS) is required for virulence of the gram-negative plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (DC3000) in tomato and *Arabidopsis*. The alternative sigma factor HrpL directly regulates expression of T3SS genes by binding to a short DNA sequence designated as the "*hrp* promoter". The ability of DC3000 to colonize plants, subdue multiple layers of plant defense and multiply in plant tissues relies on the activities carried out by the many T3SS regulon members (known collectively as *hrp* genes).

Efforts to identify genes involved in pathogenicity were initiated over three decades ago. However, HrpL binding to *hrp* promoters has never been directly demonstrated and it is unclear if the list of HrpL-regulated genes is complete. The first goal of the research described here was to systemically and exhaustively identify HrpL-binding sites and likely *hrp* promoters in the DC3000 genome. Employing chromatin immuno-precipitation, coupled with high-throughput sequencing (ChIP-Seq) and transcription start site analysis (modified RNA-Seq), we found twenty sites representing novel *hrp* promoters. Using deletion analysis, we attempted to determine if the genes downstream from a

subset of these promoters could be linked to virulence. However, the deletions did not affect the hypersensitive response or *in planta* growth of the resulting strains. Interestingly, many new HrpL regulon members appear to be unrelated to the T3SS (based on their annotations), and orthologs for some of these can be identified in non-pathogenic bacteria. The connection of these new HrpL regulon members to virulence is not obvious.

The HrpL regulon is activated as a result of a chain of events, most of which are not well understood. It is known that RpoN, which controls the transcription of *hrpL* in DC3000, is required for virulence in several bacterial species. Motivated by the hypothesis that genes are coordinately regulated in order to serve a strategic purpose (e.g., virulence), our second goal was to look for other genes activated by RpoN in parallel with *hrpL*. RpoN (σ^{54}) requires specialized enhancer-binding proteins (EBPs) in order to activate transcription. This arrangement presumably allows the cell to respond to environmental signals by modifying the transcription of particular genes. Using ChIP-Seq and RNA-Seq, we identified candidate RpoN-dependent genes as well as genes that were differentially expressed under *hrp*-inducing conditions. This initial survey includes more than 200 likely RpoN-regulated genes involved in flagella biosynthesis, energy metabolism, nitrogen metabolism, transport and binding proteins, and small noncoding RNAs, as well as putative regulatory proteins and EBPs. Among the genes that were differentially regulated between *hrp*-inducing and repressing conditions, more than one dozen appear to be regulated by RpoN

and are therefore potentially important in functions related to plant association or virulence.

BIOGRAPHICAL SKETCH

Hanh Ngoc Lam was born on October 27th, 1983 to Hung Thanh Lam and Muot Thi Nguyen in Vinh Long, a small province in the southern part of Vietnam. She is very close to her two younger sisters, whom she considers great friends. Being the oldest child in the family, Hanh has been a role model for her siblings in many aspects of life. She was passionate about science from a young age and spent hours in her own study in preparation to enter college to pursue a career in science and contribute to education.

Moving to Ho Chi Minh City for a B.S. degree in Biotechnology in 2001, Hanh was able to catch up with peers who had been raised in the city and had already been exposed to advanced classes, including English, which is often not offered in rural areas. She participated in multiple projects and eventually focused on plant viruses for her thesis research. Because of her excellent performance in classes and activities, Hanh was recruited by her own institution as an apprentice lecturer and researcher in 2005. Working professionally in a scientific and educational environment, she understood that additional experience would be essential if she wanted to pursue research and help train the next generation of scientists. She therefore began to prepare herself to study abroad.

In 2008, Hanh was awarded a fellowship from the Vietnam Education Foundation to pursue a Ph.D. in the United States. She chose to come to Cornell University to study Plant-Microbe biology. Hanh completed a rotation in Dr. Alan

Collmer's laboratory and was excited about the ways in which bacteria had evolved to combat plant disease defenses. Later on, she joined the Cartinhour lab and spent the next four years mastering experimental molecular biology and computational analysis.

Even as a youth, Hanh was always active in a variety of sports such as Taekwondo, table tennis, and long-distance running. She has maintained her interest in these activities at Cornell. She has also taken full advantage of Cornell's environment including the Slope, the Plantations and the gorges, where there are always opportunities to view or photograph scenery and campus architecture, or natural settings with leaves, flowers and birds. Cornell has been a wonderful place for her to develop good friendships and enjoy natural beauty.

To my loving family for raising me well and supporting me.

To my friends for being great company.

To my mentors for developing my mind.

ACKNOWLEDGMENTS

I would like to thank Samuel Cartinhour, who has been very patient and considerate with me. Sam would spend every Friday morning to help me to review model pathogenicity systems and plant pathology concepts to prepare for the A exam. Sam would help me with everything including English speaking and writing. My background in computational biology would not be as it is now without support from Sam. I would also like to thank everyone in the Cartinhour lab that I have had the privilege to work with, with a few special acknowledgements: to Melanie Filiatrault who taught me to do RNA-Seq and work with RNA, to Bronwyn Butcher who taught me to do ChIP-Seq, to Paul Stodgill who always gave me detailed answers when I was stuck with a programming problem, and to HoangChuong BuiNguyen who spent days and nights with me in the lab working on multiple projects. Eric Markel always gave precise answers and instructions on protocols used in the lab. Bryan Swingle gave me several plasmid constructs. Christopher Myers has been my excellent mentor in Computational Biology. John Helmann gave me valuable feedback and advice on courses and research every time we met. Keith Perry made sure that I knew and understood concepts in Plant Pathology. My work was supported by funding from the USDA-ARS (U.S. Department of Agriculture-Agricultural Research Service), by NSF funding IOS-1025642, and by a graduate research fellowship awarded to me by the Vietnam Education Foundation (VEF). I am particularly grateful to the VEF for recruiting me into their program and for providing the opportunity to study abroad.

I would like to acknowledge the Collmer lab for helping me with the HrpL regulon project. Alan Collmer provided extensive advice on prioritizing genes for further study and on key experiments. Suma Chakravarthy helped me with the translocation assay. Hai-Lei Wei provided mutants for the new effector gene and

valuable information about flagellin related mechanism.

I have had the great privilege of being friends with many students and post docs here at Cornell and Ithaca. You all are dear to me.

To my many friends and teachers in Vietnam, your phone calls and wishes provided great encouragement, and helped me continue to be the energetic and passionate person that I have ever been.

Finally I would like to thank my family for their tireless support as I worked toward my degree. You are always in my heart.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENE REGULATION IN VIRULENCE OF *PSEUDOMONAS SYRINGAE* PV. *TOMATO* DC3000

Pseudomonas has diverse life styles in different niches

Pseudomonas is a genus of Gram-negative gammaproteobacteria with 75 species [1] and a wide diversity of life styles. The pseudomonads can be found in many niches including salt and fresh water, soil, snow [2, 3], and in association with seeds [4], plant roots and foliage, and animals. The range is illustrated by saprophytes such as *Pseudomonas putida*, a common soil bacterium used for bioremediation [5-8], and *P. fluorescens*, a rhizosphere colonizing bacterium used in biological control [9, 10], to opportunistic pathogens (e.g. *P. aeruginosa*, a medically important opportunistic multikingdom pathogen [11]), to our pathogen of interest, *P. syringae* pathovar *tomato* DC3000 [12-14]. Notably, some members of the *P. syringae* group (but not DC3000) are able to induce ice nucleation at temperatures above the freezing point of water [15]. Bacteria acting as ice nuclei are thought to cause changes in the hydrological cycle in boreal regions where tundra and forests are favorable habitats of bacteria [16]. This ability may also contribute to the wide distribution of *P. syringae* in diverse and distant habitats via precipitation [3].

Many *Pseudomonas* species are significant plant and animal pathogens. *P. plecoglossicida* infects ayu, a sweetfish relative of the smelts [17, 18]. *Pseudomonas*

alcaligenes, a Gram-negative aerobic bacterium, is an opportunistic human pathogen [19]. *Pseudomonas* can infect dicots (e.g. *P. syringae* pv. *tabaci* causing wildfire disease on tobacco and bean plants [20]), or monocots (e.g. *P. oryzae*, a pathogen of rice [21]). *P. aeruginosa* is by far the best-known animal pathogen as it is responsible for nosocomial infections and is particularly dangerous to immune-compromised patients [22]. Interestingly, *P. aerugionosa* can infect both plant and animal hosts [23]. Other pseudomonads, even those ordinarily considered "benign", can also cause human disease under certain situations, such as *P. fluorescens* found in septic sacoiliitis of immunoincompetent patients [24].

Pseudomonas syringae subdues plant defenses with multiple virulence repertoires

As mentioned previously, *P. syringae* participates in the global hydrological cycle and is found on every continent in both agricultural and natural settings [3, 25]. In agricultural systems, *P. syringae* causes disease in many hosts. More than 50 *P. syringae* pathovars are collectively clustered into five major groups [26]. Overall, *P. syringae* causes various spots (*P. syringae* pv. *syringae* B728A), specks, blights (*P. syringae*. pv. *glycinea*), declines (*P. syringae* pv. *avellanae*), cankers (*P. syringae* pv. *actinidae*) and knots (*P. syringae* pv. *savastanoi*) on aerial plant parts (source: http://www.pseudomonas-syringae.org/). Each *P. syringae* strain, however, has its own narrow host range. An example is DC3000, a causal agent of bacterial speck disease on tomato and *Arabidopsis thaliana* [12-14].

The biology of plant pathogenic bacteria has been studied almost exclusively in agricultural contexts. It is now known that plant pathogenesis is an extremely complex

process that depends on the ability of the pathogen to evade or defeat the layered defenses of the host. When DC3000 lands on leaves and finds its way into the apoplast, guard cells embedded in the plant epidermis will close stomata as they detect pathogen associated molecular patterns (PAMPs) such as lipopolysaccharides and peptides derived from bacterial flagellin. This system is not completely secure because some stomata do not respond and remain open, providing routes for low levels of bacterial entry [27]. Once in the apoplast, DC3000 secretes coronatine, a jasmonate-isoleucine mimic compound, to promote stomatal re-opening [28, 29], supporting the entry of additional bacterial cells to the apoplastic space. Coronatine also affects the development of chlorotic and necrotic lesions later in the infection by inhibiting the accumulation of the key plant immune signal salicylic acid (SA) [30, 31] (as salicylic acid and its cognate derivative are signal molecules required for systemic resistance [32, 33]). Although the apoplast is potentially a favorable environment for *P. syringae* since it offers amino acids and perhaps other nutrients [34], receptors on plant cell membrane can detect the presence of PAMPs, particularly flagellin, and trigger the innate immune response [35, 36]. The bacteria, in response, inject effector proteins through the Type III secretion system (T3SS) to suppress plant defenses in order to survive and multiply [31, 37-40].

This thesis contributes new information concerning the regulation of bacterial genes that occurs in response to environmental changes, and focuses specifically on systems that are related to the infection process. Chapter 2 presents a thorough inventory of genes directly regulated by sigma factor HrpL, best known as a direct regulator of T3SS. Among these newly identified candidates are PSPTO_5633 (*hopBM1*), a novel effector, as well as genes not obviously related to the T3SS. This second category suggests that our

view of HrpL may need revision to include its potential role in regulating functions other than those immediately involved in infection (e.g., other kinds of plant interactions). Chapter 3 considers the possibility that the sigma factor RpoN, which directs the transcription of *hrpL*, may control other genes in parallel with *hrpL* in order to support additional pathogenesis-related functions. RpoN is a high-level regulator that controls operons involved in processes ranging from metabolism to motility, virulence and the stress response. The versatility is due largely to the fact that this sigma factor requires the presence of a bacterial enhancer binding protein in order to initiate transcription. Finally, chapter 4 offers an outlook on some of the important questions that remain to be addressed experimentally.

The remainder of this introduction will cover basic knowledge of Type III secretion system, the role of sigma factors in gene regulation in DC3000, regulation of the T3SS, and the HrpL and RpoN regulons.

Type III secretion system

The T3SS consists of structural and nonstructural components [41]. The structural proteins are assembled to create a membrane-embedded basal structure, including a needle-like component protruding from the bacterial surface. This structure shares considerable homology with the flagella system that is used for motility [42, 43]. Although the mechanistic details of needle penetration and function are not yet understood, it is thought that a special structure known as the translocon bridges the needle tip and host cell membrane and creates a pore for effectors to travel from the bacterial to host

cytoplasm. In the *P. syringae* chromosome, the structural genes are grouped in *hrp*/*hrc* clusters which are conserved among closely related *Pseudomonas* strains, while the effector genes are located at the conserved effector locus (CEL), exchangeable effector locus (EEL) and elsewhere in the genome [44].

Clusters of T3SS genes are presumed to have been acquired by horizontal gene transfer from bacteria sharing the same habitats [45]. These virulence determinant clusters are located in pathogenicity islands, flanked by mobile genetic elements, and are maintained by increasing selection fitness for the acquired strains [44, 46, 47]. Thus comparative genomics, combined with the identification of mobile elements and horizontally acquired regions, provides a roadmap to identification of candidate virulence genes [48].

Although the T3SS has been studied for decades because of its importance to virulence in many gram-negative T3SS$^+$ bacteria, our understanding of virulence mechanisms and interactions between *P. syringae* and its hosts is still limited. GacA, which controls expression of *hrpL* via the sigma factor RpoN, acts as a high-level regulator of virulence in *P. syringae* and consequently plays a crucial role in the deployment of the T3SS in DC3000 and *P. syringae* pv. *phaseolicola* [49, 50]. One might expect a *gacA*$^-$ strain to be deficient in virulence, i.e., to exhibit dampened disease symptoms and reduced growth upon contact with hosts due to the lack of a functional T3SS. Surprisingly, a *gacA* mutant in a closely related strain, *P. syringae* pv. *syringae* B728A, can grow as well *in planta* as the *gacA*$^+$ strain without causing lesions [51]. In addition, a T3SS$^-$ *P. syringae* strain from a non-agricultural habitat (designated as TA-002), lacking *hopI1*, *avrE*, *hrpW* (the CEL), *hrpK*, *hrpL* and *hrcC* (*hrp*/*hrc* clusters), can still cause disease on tobacco plants [52].

Finally, the T3SS exists in non-pathogenic bacteria as well. In *rhizobia*, a nitrogen fixing genus, the T3SS is dispensable for nodulation and can affect mutualistic associations with a host [53] either positively or negatively. These variations suggest that the T3SS is ancient and its role is not only restricted to parasitism but also includes more general interactions between bacteria and hosts as mutualism and commensalism [54].

Gene regulation in DC3000: the role of sigma factors

Bacteria have a remarkable ability to adapt to rapidly changing environments by adjusting their gene expression profiles. Bacterial gene regulation is often regulated at the transcriptional level by the multi-subunit DNA-dependent RNA polymerase enzyme complex. The complex includes components that are always, or nearly always present (α subunits forming dimers with which the β and β' subunits associate [55]) as well as components that are transiently associated (such as sigma factors). The sigma factor subunit plays an important role in selecting which genes can be transcribed through its ability to localize the polymerase upstream of a gene prior to the initialization of mRNA synthesis. This function depends on the recognition by the sigma factor of a promoter, a short DNA sequence that essentially serves as a docking site. Generally, transcription begins less than 10 bp downstream from the promoter element [56].

In rapidly growing bacteria, RpoD (often generically designated σ^{70} due to its similarity to σ^{70} of *Escherichia coli*) is the primary sigma factor responsible for transcription of most "housekeeping" genes such as those involved in central metabolism. However, under certain conditions, a different subset of genes may be deployed (for example, if the cell is stressed due to nutrient limitation or other environmental changes, or

if the kind of nutrients that are available in the medium changes). The change in gene expression is sometimes supported by alternative sigma factors, allowing the cell to adjust its metabolic capacity to adapt to a different environment [57]. Alternative sigma factors include members in the σ^{70} family, similar but not identical to σ^{70} of *E. coli,* and also the σ^{54} family [57, 58]. Within the σ^{70} family, sigma factors can be further classified into four phylogenetic subgroups, among which the extracytoplasmic function (ECF) σ factors are the largest and most diverse group [59]. HrpL is the most prominent member of the ECF family.

Pseudomonas syringae pv. *tomato* DC3000 has genes encoding 15 sigma factors [60] including 10 ECF sigma factors [61]. Global transcription start site mapping of late exponential phase bacteria in moderate *hrp*-inducing conditions (MG medium) made it possible to identify promoter motifs closely resembling known promoters for six sigma factors including HrpL (regulator of the T3SS), RpoN (σ^{54}), RpoF (regulator of flagella biosynthesis), RpoE (AlgU), RpoH (regulator of heat shock genes), and RpoD (σ^{70}) [62]. Regulons controlled by five iron starvation sigma factors have also been characterized with functions involved in the production or uptake of pyoverdine (by PvdS) [63], expression of genes in response to hydroxamate siderophore (by PSPTO_1203) [64], expression of iron-regulated membrane protein (by PSPTO_0444), expression of heme oxygenase (by PSPTO_1286), and genes of unknown function regulated by PSPTO_1209 [60]. In DC3000, HrpL is without a doubt the most studied sigma factor because of its involvement in the T3SS. The HrpL regulon has been globally investigated using microarrays, revealing 195 differentially expressed genes dependent on HrpL [65], including genes regulated by HrpL both directly and indirectly.

Regulation of T3SS in *Pseudomonas syringae*

Activation of the T3SS in *P. syringae* is initiated by the two-component system GacS/GacA. GacS (originally LemA), is a sensory histidine kinase that transduces signals to GacA, which is a response regulator. This system regulates many bacterial phenotypes including stress tolerance and virulence [66, 67]. In DC3000, GacS/GacA affects the transcription of *rpoN* and *hrpRS* [49] although the signals and mechanism of regulation are unknown. RpoN, in turn, regulates expression of *hrpL* [68] in the presence of the EBPs HrpR and HrpS, which form heterodimers binding to a regulatory region upstream of *hrpL* [69, 70]. HrpL then supports the transcription of T3SS genes via its recognition of the *hrp* promoter [71]. Figure 1.1, modified from [72], depicts the regulation of T3SS.

There are positive and negative regulators of the T3SS. HrpA1 (a type III helper protein) is required for full expression of other Hrp proteins such as HrpW, HrcJ and HrcC. Experiments involving the deletion of *hrpA1* showed that it affects expression of *hrp* genes at the level of transcription [73]. HrpA1 further helps secretion of other effectors including AvrPto [73]. It is unknown whether HrpA1 directly or indirectly regulates the level of HrpRS. The negative regulation of T3SS involves Lon protease [74], HrpG and HrpV [75]. Lon-mediated degradation maintains HrpR at the basal level in *hrp*-repressing medium (nutrient-rich medium). Elevated expression of *hrpR* in *hrp*-inducing conditions brings the level of HrpR to the concentration required for it to activate the T3SS [74]. The activity of HrpS, which interacts with HrpR, is inhibited by direct binding of HrpV. Finally, HrpG can bind to HrpV, acting as an anti-negative regulator of T3SS [75].

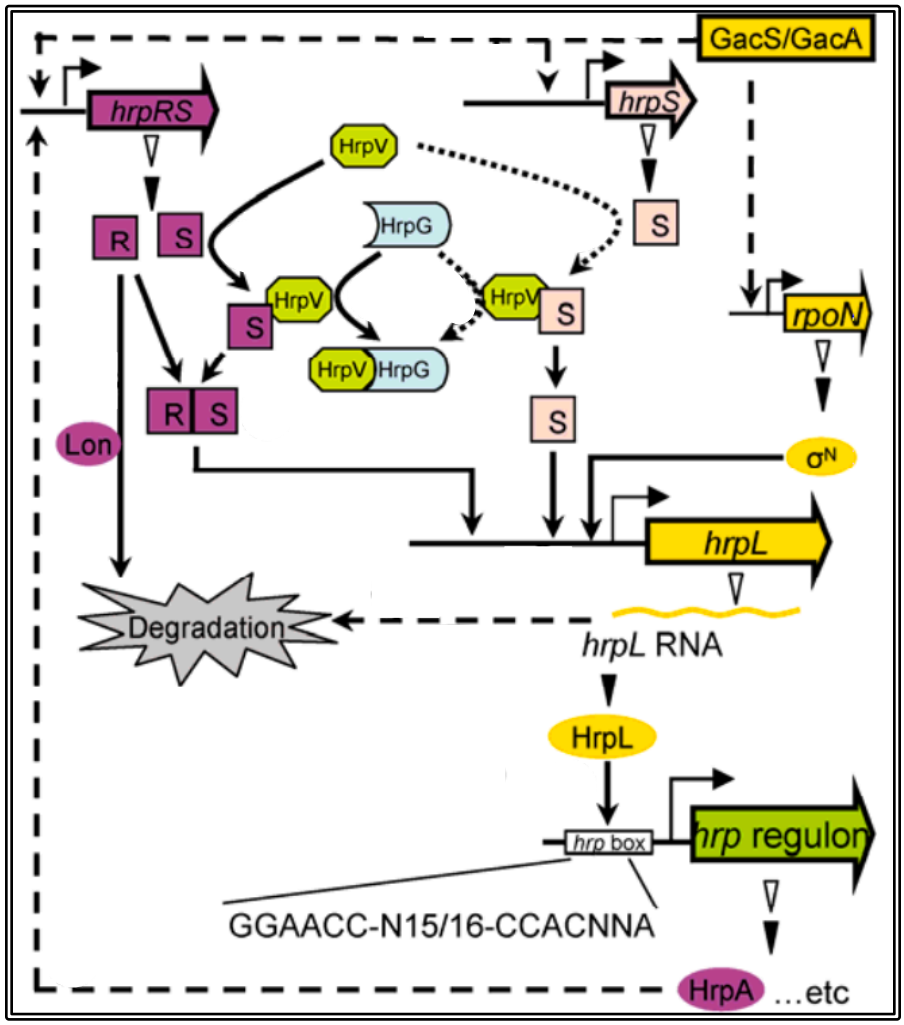


Figure 1. 1 Regulation of Type III secretion system in *Pseudomonas syringae*

Figure is modified from reference [72]. Expression of the T3SS is regulated by the two-component system GacS/GacA, HrpA1 (helper protein), HrpS/HrpR (EBPs), sigma factor RpoN and Lon protease.

HrpL regulon

The characterization of *hrp* regulon members began in the 1980's with the identification of genes responsible for an "avirulence" phenotype, so named because it is the result of a bacterial protein or its activity being recognized and neutralized by the plant immune system [76-78]. Subsequently, the HrpL sigma factor and the promoters it recognizes that together are responsible for the transcription of avirulence genes were confirmed [71]. This foundation made it feasible to conduct a more systematic search for HrpL promoter elements using a bioinformatics-based approach and high-throughput screening [65, 79-85]. These efforts resulted in a list of HrpL regulated genes and operons thought to be nearly complete at the time [86-88].

Chapter 2 addresses the question of the completeness of the HrpL regulon and presents multiple lines of evidence supporting the existence of new HrpL-regulon members. The microarray technology that was originally used to establish the most-accepted list of HrpL regulon members is probably not as sensitive as RNA-seq or other more recent high-throughput sequencing methods. This was foreshadowed by the fact that some genes were shown to be differentially expressed using RT-PCR even though their levels were unchanged in the microarray experiment [65]. Critically, without evidence for direct binding of *hrp* promoters to HrpL, some potential *hrp* promoters remained unconfirmed because their sequences were divergent from the consensus sequence or far from the downstream genes. The increasingly popular approach to studying pathogenesis using minimal functional repertoire strains, from which all active effectors must be removed, also motivates interest in a complete list of virulence genes and effectors [89, 90].

Chapter 2 describes experiments in which we confirm 20 new HrpL-responsive elements, providing valuable information for future studies of bacterial pathogenesis.

RpoN regulon

RpoN was initially identified as a gene involved in Nitrogen fixation [91]. It is widespread in bacteria, and its target genes carry out a wide range of different functions that are not obviously related [92]. RpoN is distinct from σ^{70} both structurally and functionally, and is unique among bacterial sigma factors in its mode of transcription initiation [93]. Another distinct feature of RpoN is that it can bind to promoter DNA in the absence of core RNA polymerase [94], while most members of the sigma 70 family cannot bind to DNA in the absence of RNA polymerase [95, 96].

RpoN and RpoN-RNA polymerase holoenzyme (RpoN-RNAP) has properties reminiscent of eukaryotic RNA polymerase II. RpoN has several functional motifs normally found in eukaryotic transcription factors such as a leucine zipper and glutamine-rich region [97]. RpoN-RNAP can be activated at a distance by EBPs binding to enhancer-like elements (ELE), and requires nucleotide hydrolysis catalysed by EBPs for transcription initiation to proceed [98].

Enhancer-like elements (ELEs) are palindromic or nearly palindromic nucleotide sequences found 100 bps or more upstream of certain group of promoters in prokaryotes. Remarkably, like eukaryote enhancers, they can be functional even when moved several kilobases upstream or downstream away from the target genes regardless of their orientation [99, 100]. ELEs are binding sites for regulatory proteins known generically as enhancer binding proteins, or EBPs. There are, however, certain strong promoters whose transcription can be activated by EBPs without involvement of ELEs, although both the

advantages conferred by this arrangement and the mechanism by which it proceeds remain to be studied [101].

The requirement for an EBP for transcriptional activation of RpoN regulated genes was first noted for the glutamine synthetase gene of *E. coli* [100]. The unique requirement for an EBP allows bacteria to integrate environmental signals with gene regulation via enhancer availibility. Enhancer activation signals have been difficult to identify. In some cases, small molecule ligands can activate EBPs [102].

Three factors are usually required to activate transcription for RpoN dependent genes, namely, integration host factor (IHF) [103, 104], EBPs [105, 106] and adenosine triphosphate (ATP) [107]. IHF facilitates DNA bending to bring EBPs, which are typically bound at a distance from the RpoN binding site, into proximity with the σ^{54}-closed complex (RNA polymerase holoenzyme). EBPs initially associate with σ^{54} weakly. But after ATP is bound to the EBP center domain (via the AAA+ domain involved in ATP binding and hydrolysis), the EBP conformation changes and a stronger association with σ^{54} is established. This repositions σ^{54} and the holoenzyme-closed DNA complex, facilitating transcription initiation [108].

RpoN is a master regulator in many different bacteria [109]. It regulates genes involved in structures and functions such as the type VI secretion system [110], type IV pilus [104], *in planta* growth of bacteria and the provocation of the plant hypersensitive response [111], virulence, and the formation of biofilms [112]. The RpoN regulon has been studied in many prokaryotes including the gram-negative sulfur-reducing bacterium *Geobacter sulfurreducens* [113], the soil bacterium *Rhizobium etli* [114], the cholera causal

agent *Vibrio cholera* [115], a microsymbiont *Mesorhizobium loti* [116], and the human pathogen *Pseudomonas aeruginosa* [117].

In DC3000, RpoN is responsible for transcription of *hrpL* [49]. However it is unknown whether RpoN directly regulates other virulence related genes or any of the processes with which it is associated in other organisms. Chapter 3 will review the current understanding of RpoN and its regulon members, and presents a preliminary inventory of the RpoN regulon using ChIP-Seq, RNA-Seq and computational analyses. 226 RpoN binding sites were found (associated with 214 RpoN motifs). 12 genes potentially regulated by RpoN are differentially expressed. RpoN regulated genes in DC3000 appear to support diverse functions including flagella biosynthesis, ligand binding and membrane transport, small noncoding RNAs, and regulators. These will be valuable targets for further study by those studying gene regulation, bacterial pathogenesis and many other physiological mechanisms in DC3000.

CHAPTER 2

GLOBAL ANALYSIS OF THE HRPL REGULON IN THE PLANT PATHOGEN *PSEUDOMONAS SYRINGAE* PV. *TOMATO* DC3000 REVEALS NEW REGULON MEMBERS WITH DIVERSE FUNCTIONS [*]

Abstract

The type III secretion system (T3SS) is required for virulence in the gram-negative plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000. The alternative sigma factor HrpL directly regulates expression of T3SS genes via a promoter sequence, often designated as the "*hrp* promoter." Although the HrpL regulon has been extensively investigated in DC3000, it is not known whether additional regulon members remain to be found. To systematically search for HrpL-regulated genes, we used chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) and bulk mRNA sequencing (RNA-Seq) to identify HrpL-binding sites and likely *hrp* promoters. The analysis recovered 73 sites of interest, including 20 sites that represent new *hrp* promoters. The new promoters lie upstream of a diverse set of genes encoding potential regulators, enzymes and hypothetical proteins. PSPTO_5633 is the only new HrpL regulon member that is potentially an effector and is now designated HopBM1. Deletions in several other new regulon members, including PSPTO_5633, PSPTO_0371, PSPTO_2130, PSPTO_2691, PSPTO_2696, PSPTO_3331, and PSPTO_5240, in either DC3000 or Δ*hopQ1-1*

[*] Hanh N. Lam, Suma Chakravarthy, Hai-Lei Wei, HoangChuong BuiNguyen, Paul V. Stodghill, Alan Collmer, Bryan M. Swingle, and Samuel W. Cartinhour. Global analysis of the HrpL regulon in the plant pathogen

backgrounds, do not affect the hypersensitive response or *in planta* growth of the resulting strains. Many new HrpL regulon members appear to be unrelated to the T3SS, and orthologs for some of these can be identified in numerous non-pathogenic bacteria. With the identification of 20 new *hrp* promoters, the list of HrpL regulon members is approaching saturation and most likely includes all DC3000 effectors.

Introduction

Pseudomonas syringae pv. *tomato* DC3000 (DC3000), an important model pathogen in molecular plant pathology, causes bacterial speck disease in *Arabidopsis* [13], tomato [14] and *Nicotiana benthamiana* (DC3000 mutants lacking virulence determinant HopQ1-1) [118]. The ability of DC3000 to colonize plants and subdue multiple layers of plant defense is dependent on the type III secretion system (T3SS) [119]. The T3SS machinery is encoded by the hypersensitive response and pathogenicity (*hrp*) and *hrp* conserved (*hrc*) gene clusters [120]. Effector proteins, encoded by *hrp* outer protein genes (*hop*) [121] are translocated into the host cytoplasm via the T3SS to interact with host proteins and/or intervene with host signaling cascades and responses for the benefit of the pathogen [122-124]. However, if one or more effectors or its activities are recognized by the plant immune system (through resistance proteins or other mechanisms), the host hypersensitive response (HR), a localized plant cell death, is triggered and bacterial growth is limited [125]. Effectors are examples of avirulence (avr) genes, a diverse group whose products typically stimulate a strong host defense response.

Efforts to identify genes involved in virulence and pathogenicity were initiated well before the genome sequences of DC3000 and other *P. syringae* strains were available.

Experimental approaches included screening cosmid libraries for gain-of-function avirulence phenotypes [76, 77], using partial sequencing to characterize gene clusters flanked by pathogenicity islands [126], and identifying proteins secreted by the T3SS [81, 83, 127]. Regulation of the T3SS was linked to the alternative sigma factor HrpL when a DNA sequence of length 32 bases upstream of *hrpZ* was recognized to support HrpL-dependent transcription [71, 78, 128]. More comprehensive analyses of the HrpL regulon were possible once the genome sequence was determined [129]. For example, a combination of promoter trapping and sequence analysis was used to identify functional *hrp* promoters associated with pathogenicity [80]. This work established a consensus sequence for the putative *hrp* promoter. Another high-throughput screen, also based on promoter trapping, identified 29 T3SS proteins in DC3000 [85]. Although the search for T3SS effectors was suggested to be near saturation in this screen, it was carried out using a HrpL overexpression system which might have been vulnerable to false positives. A microarray screen comparing WT DC3000 and a Δ*hrpL* mutant [65] generated the currently accepted list of HrpL regulon members and 54 annotated *hrp* promoters, and has been considered to be nearly complete or exhaustive [86, 87]. More recently, HrpL overexpression and RNA-Seq were used to analyze the HrpL regulon in DC3000 as well as five other *P. syringae* strains, resulting in the identification of 14 new potential regulon members in DC3000 [130].

Although multiple approaches have been used to characterize the DC3000 HrpL regulon, several factors suggest that additional members may yet be found. First, inventory strategies have emphasized the identification and functional characterization of effectors, and thus non-effector genes that are important to bacterial growth and survival may have

been overlooked. An example of a gene in this class is PSPTO_0834 (alcohol dehydrogenase, zinc-containing protein) [88], which is regulated by HrpL and strongly influences bacterial virulence. Second, effector identification is challenging because of effector redundancy and the frequent failure of effector gene mutants to exhibit a virulence phenotype. Many effectors, moreover, show no similarity to known proteins. Third, the *in planta* growth of D28E, a T3SS^{+} DC3000 derivative from which the 18 clustered effectors and 10 additional effectors were deleted, is significantly reduced relative to a T3SS^{-} control strain [89, 90]. This suggests that D28E expresses as yet unidentified factors that are recognized by the plant, some of which may be HrpL regulon members. Fourth, it is not yet known whether divergent examples of the *hrp* promoter, such as the one upstream of *iaaL* [65] actually bind HrpL or promote HrpL-dependent transcription. Finally, newly available genomic technologies offer enhanced sensitivity for the detection of transcriptional activity and promoter identification [130-132]. Together, these factors indicate that reexamining the HrpL regulon in DC3000 for additional members would be fruitful.

Here, we use chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq), and a modified RNA-Seq protocol (incorporating the capture of mRNA 5'-ends), to provide evidence for direct binding of HrpL at *hrp* promoters and their activation by HrpL. The experiments identified most known members of the regulon (representing 52 out of 54 previously annotated HrpL-dependent promoters [65, 80, 81, 83, 85-87, 133]), as well as 20 new *hrp* promoter candidates. HrpL binding was validated using ChIP followed by quantitative polymerase chain reaction analysis (ChIP-qPCR) and promoter activity was tested using transcriptional fusions and quantitative reverse transcription-PCR (qRT-PCR). Computational analyses were used to search for additional

members of the HrpL regulon in DC3000 and to investigate the conservation of HrpL regulon members and potential *hrp* promoters within the available *Pseudomonadales* genomes. This analysis revealed that some new HrpL regulon members are widely represented in both pathogenic and non-pathogenic bacteria. Finally, we conducted a translocation assay to demonstrate that one new HrpL regulon member, PSPTO_5633 (designated HopBM1), is translocated into plant cells in a T3SS-dependent manner.

Materials and methods

Bacterial strains and growth conditions

DC3000 and its derivatives were grown at 28°C in MG medium [134], Kings B (KB) medium [135], HMM (*hrp*-minimal medium) [136], or on KB, MG, and HMM media solidified with 1.5% (wt/vol) agar. *Escherichia coli* TOP10 (Invitrogen) was used as the host for sub-cloning and other plasmid manipulations. *E. coli* was grown at 37°C in Luria-Bertani (LB) [137] medium or LB medium solidified with 1.5% (wt/vol) agar. Rifampicin, spectinomycin, kanamycin, and tetracycline were used at 50 μg/ml, 50 μg/ml, 50 μg/ml and 10 μg/ml, respectively. Rifampicin was added to medium used for growth of DC3000 derivatives. Other antibiotics were added to maintain plasmids as indicated in Table 2.1.

Table 2. 1 All strains used in this study

Designation	Genotype and Relevant Features	Reference
Plasmids		
pK18mobsacB	Small mobilizable suicide vector, sucrose-sensitive (sacB)/ Kan^R	[138]
pHL1	Destination vector for transcriptional fusion using LR reaction, carrying *gfp*/ Tet^R Cam^R	This study
pBS181	p*hrpJ*::iucD expresses GUS under T3SS inducing conditions/ Tet^R Kan^R	This study
pCPP6424	pENTR/SD/D-TOPO::PSPTO_5633; Entry vector with the coding region of PSPTO_5633 lacking the stop codon/ Kan^R	This study
pCPP5371	Cya Gateway destination vector/ Gen^R Cam^R	[139]
pCPP6413	pCPP5371::PSPTO_5633/ Gen^R	This study
pCPP5388	pCPP5371::AvrPto/ Gen^R	This study
pBS62	pENTR/D with DC3000 genomic coordinates 1542594..1542860; contains *hrpJ* promoter/ Kan^R	This study
pBS180	*iucD* reporter promoter trap destination vector; Gateway cassette cloned upstream of *iucD*/ Tet^R Kan^R Cam^R	This study
pBS59	*lux* operon reporter promoter trap destination vector; Gateway cassette cloned upstream of *lux* operon/ Tet^R Kan^R Cam^R	[63]
pTGS	Source of *gfp mut2*/Tet^R	[140]
pBS12	*gfp* reporter promoter trap vector/ Tet^R	This study
HLN068	pENTR/SD/D-TOPO::P_{PSPTO_0871}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN069	pENTR/SD/D-TOPO::P_{PSPTO_0816}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN100	pENTR/SD/D-TOPO::P_{PSPTO_2130}; Entry vector containing candidate *hrp*	This study

	promoter upstream from noted gene/ Kan^R	
HLN101	pENTR/SD/D-TOPO::P_{PSPTO_4750}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN102	pENTR/SD/D-TOPO::P_{PSPTO_5053}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN103	pENTR/SD/D-TOPO::P_{PSPTO_5618}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN105	pENTR/SD/D-TOPO::P_{PSPTO_1843}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN106	pENTR/SD/D-TOPO::P_{PSPTO_3481}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN107	pENTR/SD/D-TOPO::P_{PSPTO_3721}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN108	pENTR/SD/D-TOPO::P_{PSPTO_4340}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN206	pENTR/SD/D-TOPO::P_{PSPTO_5633}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN207	pENTR/SD/D-TOPO::P_{PSPTO_0371}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN208	pENTR/SD/D-TOPO::P_{PSPTO_1645}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN210	pENTR/SD/D-TOPO::P_{PSPTO_2691}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN212	pENTR/SD/D-TOPO::P_{PSPTO_2696}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN214	pENTR/SD/D-TOPO::P_{PSPTO_3331}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study

HLN215	pENTR/SD/D-TOPO::$P_{PSPTO\_3948\_9}$; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN216	pENTR/SD/D-TOPO::P_{PSPTO_4699}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN217	pENTR/SD/D-TOPO::P_{PSPTO_4721}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN218	pENTR/SD/D-TOPO::P_{PSPTO_4955}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN250	pENTR/SD/D-TOPO::P_{PSPTO_5240}; Entry vector containing candidate *hrp* promoter upstream from noted gene/ Kan^R	This study
HLN072	pHL1::P_{PSPTO_0871}::*gfp*; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN162	pHL1::P_{PSPTO_2130}::*gfp*; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN163	pHL1::P_{PSPTO_4750}::*gfp*; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN164	pHL1::P_{PSPTO_5053}::*gfp*; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN165	pHL1::P_{PSPTO_5618}::*gfp*; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN167	pHL1::P_{PSPTO_1843}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN168	pHL1::P_{PSPTO_3481}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN169	pHL1::P_{PSPTO_3721}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study

HLN170	pHL1::P_{PSPTO_4340}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN219	pHL1::P_{PSPTO_5633}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN220	pHL1::P_{PSPTO_0371}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN221	pHL1::P_{PSPTO_1645}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN223	pHL1::P_{PSPTO_2691}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN225	pHL1::P_{PSPTO_2696}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN227	pHL1::P_{PSPTO_3331}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN228	pHL1::$P_{PSPTO\_3948\_9}$; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN229	pHL1::P_{PSPTO_4699}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN230	pHL1::P_{PSPTO_4721}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN231	pHL1::P_{PSPTO_4955}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN253	pHL1::P_{PSPTO_0816}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study
HLN254	pHL1::P_{PSPTO_5240}; Expression vector containing candidate *hrp* promoter fused to *gfp* reporter/Tet^R	This study

Escherichia coli		
TOP10	Φ80*lacZΔM15 ΔlacX74*	Invitrogen
Pseudomonas syringae pv. tomato		
DC3000	DC3000/ Rif^R	[141]
CUCPB5460	DC3000 with *hopQ1-1* deleted (*ΔhopQ1-1*)/Rif^R	[118]
UNL-134-1	DC3000 with *hrpL* deleted (Δ*hrpL*)/Rif^R Spc^R	[65]
DC3000pBS181	DC3000 carrying p*hrpJ*::iucD; expresses GUS under T3SS inducing conditions/Tet^R Kan^R Rif^R	This study
*hrpL-FLAG*pBS181	HLN090 carrying p*hrpJ*::iucD; expresses GUS under T3SS inducing conditions/Tet^R Kan^R Rif^R	This study
HLN090	DC3000 with *hrpL* tagged by FLAG at its C-terminus (*hrpL-FLAG*)/Rif^R	This study
HLN010	DC3000 with *PSPTO_5633* on chromosome deleted; remaining another copy on plasmid B/ Rif^R	This study
HLN009	DC3000 with hopQ1-1 and *PSPTO_5633* on chromosome deleted/ Rif^R	This study
HLN330	DC3000 with *PSPTO_5633* on plasmid and chromosome deleted/ Rif^R	This study
HLN012	DC3000 with *PSPTO_0371* deleted/ Rif^R	This study
HLN011	DC3000 with *hopQ1-1* and *PSPTO_0371* deleted/ Rif^R	This study
HLN014	DC3000 with *PSPTO_2130* deleted/ Rif^R	This study
HLN013	DC3000 with *hopQ1-1* and *PSPTO_2130* deleted/ Rif^R	This study
HLN016	DC3000 with *PSPTO_2691* deleted/ Rif^R	This study
HLN015	DC3000 with *hopQ1-1* and *PSPTO_2691* deleted/ Rif^R	This study
HLN018	DC3000 with *PSPTO_2696* deleted/ Rif^R	This study
HLN017	DC3000 with *hopQ1-1* and *PSPTO_2696* deleted/ Rif^R	This study
HLN020	DC3000 with *PSPTO_3331* deleted/ Rif^R	This study

HLN019	DC3000 with *hopQ1-1* and *PSPTO_3331* deleted/ Rif^R	This study
HLN182	DC3000 with *PSPTO_5240* deleted/ Rif^R	This study
HLN183	DC3000 with *hopQ1-1* and *PSPTO_5240* deleted/ Rif^R	This study
HLN190	Δ*pvsA* carrying pHL1::P_{PSPTO_2130}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_2130*/Rif^R Tet^R Kan^R	This study
HLN191	ΔpvsA carrying pHL1::P_{PSPTO_4750}::*gfp*; *gfp* under control of promoter antisense of *PSPTO_4750*/Rif^R Tet^R Kan^R	This study
HLN192	ΔpvsA carrying pHL1::P_{PSPTO_5053}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_5053*/Rif^R Tet^R Kan^R	This study
HLN193	ΔpvsA carrying pHL1::P_{PSPTO_5618}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_5618*/Rif^R Tet^R Kan^R	This study
HLN195	ΔpvsA carrying pHL1::P_{PSPTO_1843}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_1843*/Rif^R Tet^R Kan^R	This study
HLN196	ΔpvsA carrying pHL1::P_{PSPTO_3481}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_3481*/Rif^R Tet^R Kan^R	This study
HLN197	ΔpvsA carrying pHL1::P_{PSPTO_3720}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_3720*/Rif^R Tet^R Kan^R	This study
HLN198	ΔpvsA carrying pHL1::P_{PSPTO_4340}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_4340*/Rif^R Tet^R Kan^R	This study
HLN232	ΔpvsA carrying pHL1::P_{PSPTO_5633}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_5633*/Rif^R Tet^R Kan^R	This study
HLN233	ΔpvsA carrying pHL1::P_{PSPTO_0371}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_0371*/Rif^R Tet^R Kan^R	This study
HLN234	ΔpvsA carrying pHL1::P_{PSPTO_1645}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_1645*/Rif^R Tet^R Kan^R	This study
HLN237	ΔpvsA carrying pHL1::P_{PSPTO_2691}::*gfp*; *gfp* under control of promoter	This study

	upstream of *PSPTO_2691*/Rif^R Tet^R Kan^R	
HLN238	ΔpvsA carrying pHL1::P_{PSPTO_2696}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_2696*/Rif^R Tet^R Kan^R	This study
HLN240	ΔpvsA carrying pHL1::P_{PSPTO_3331}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_3331*/Rif^R Tet^R Kan^R	This study
HLN241	ΔpvsA carrying pHL1::$P_{PSPTO\_3948\_3949}$::*gfp*; *gfp* under control of a *hrp* promoter in the middle *PSPTO_3948* and *PSPTO_3949*/Rif^R Tet^R Kan^R	This study
HLN242	ΔpvsA carrying pHL1::P_{PSPTO_4699}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_4699*/Rif^R Tet^R Kan^R	This study
HLN243	ΔpvsA carrying pHL1::P_{PSPTO_4721}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_4721*/Rif^R Tet^R Kan^R	This study
HLN244	ΔpvsA carrying pHL1::P_{PSPTO_4955}::*gfp*; *gfp* under control of promoter upstream of *PSPTO_4955*/Rif^R Tet^R Kan^R	This study
HLN263	ΔpvsA carrying pHL1::P_{PSPTO_0816}::*gfp*; *gfp* under control of a promoter upstream of of *PSPTO_0816*/Rif^R Tet^R Kan^R	This study
HLN264	ΔpvsA carrying pHL1::P_{PSPTO_0871}::*gfp*; *gfp* under control of a promoter upstream of *PSPTO_0871*/Rif^R Tet^R Kan^R	This study
HLN265	ΔpvsA carrying pHL1::P_{PSPTO_5240}::*gfp*; *gfp* under control of a promoter upstream of *PSPTO_5240*/Rif^R Tet^R Kan^R	This study
HLN245	ΔpvsA carrying pHL1::P_{ϕ}::*gfp*; ; empty plasmid as negative control in *gfp* assay/Rif^R Tet^R Kan^R	This study
PS167	DC3000 *ΔpvsA*/ Rif^R Kan^R	[142]
DC3000T3SS-	CUCPB5113, DC3000*ΔhrcQ-U*/ Rif^R Spc^R	[143]
DC3000T2SS-	DC3000*ΔgspD*/ Spc^R	[144]
HLN397	DC3000 T3SS- carrying PSPTO_5633-Cya fusion (pCPP6413)/Rif^R Gen^R	This study
HLN078	DC3000 carrying PSPTO_5633-Cya fusion (pCPP6413)/Rif^R Gen^R	This study

HLN396	DC3000 T2SS- carrying PSPTO_5633-Cya fusion (pCPP6413)/Rif^R Gen^R	This study
HLN079	DC3000 carrying AvrPto-Cya fusion (pCPP5388)/Rif^R Gen^R	This study
HLN074	DC3000 T3SS- carrying AvrPto-Cya fusion (pCPP5388)/Rif^R Gen^R	This study

Routine bacterial growth and medium shift experiments from *hrp*-inhibiting (KB medium) to *hrp*-inducing (MG supplemented with ferric citrate) conditions were carried out as follows. Colonies of *hrpL-FLAG* (in which the *hrpL* gene at its native locus has been tagged with a C-terminal FLAG epitope; see below for construction) and $\Delta hrpL$ were obtained from KB plates that had been incubated for 48 hours. Cells were re-suspended and grown overnight in 250 ml KB at 28°C with shaking at 250 rpm. Cultures were pelleted by centrifugation and washed in MG. Washed cells were re-suspended in MG and inoculated into bioreactors (Infors-HT, Switzerland) containing 400 ml MG medium supplemented with ferric citrate (Sigma-Aldrich) to a final concentration of 50 μM as described previously [134]. Samples were collected for RNA-Seq (5 ml) and ChIP-Seq (100 ml) at 1.5 hours after the medium shift into MG. Samples for RNA-Seq were supplemented with two volumes of RNAprotect Bacteria (Qiagen) to stabilize RNAs and stored at -70^0C before RNA extraction. Samples for ChIP-Seq were immediately cross-linked with 37% Formaldehyde (1% final concentration) for 20 minutes. The crosslinking reaction was quenched with 2.5 M Glycine (0.36 M final concentration). Cells were collected and washed twice with Tris-buffered saline (TBS). Washed pellets were stored at -70^0C until further processing.

Construction of plasmids and strains

Suicide vectors for gene deletions, single-crossover insertions or other purposes were introduced into DC3000 backgrounds using electroporation [145]. Deletions created using pK18mobsacB (lacking FLP recombination target (FRT) cassettes) were performed as described previously [146]. Plasmid insertions into the bacterial chromosome were

selected by plating on KB medium with kanamycin. Plasmid integration was confirmed by PCR, antibiotic resistance and sequencing.

Construction of hrpL-FLAG. Regions flanking the PSPTO_1404 *(hrpL)* gene were amplified by primers oSWC04110/oSWC04112 and oSWC04114/oSWC04116 (see Table 2.2 for all primer sequences) from DC3000 genomic DNA, purified by gel electrophoresis and gel extraction (Qiagen), and joined by SOEing PCR [147]. The joined fragment was then digested with XbaI (all enzymes were obtained from New England Biolabs unless otherwise noted), ligated with XbaI digested pK18mobsacB, and then transformed into *E. coli* TOP10. To ensure that the resulting construct was free from unwanted mutations, it was sequenced using primers M13F, M13R, oSWC05110, oSWC05111, oSWC05112, and oSWC05113. The FLAG-tagged construct was introduced into DC3000 by electroporation to generate HLN090. Merodiploid intermediates were selected for growth on medium containing kanamycin. Recombinants that had eliminated pK18mobsacB plasmid sequences were identified by sucrose counter-selection. Sucrose-resistant, Kan-sensitive colonies were analyzed by Sanger sequencing (with primers oSWC04110, oSWC04112, oSWC04114, oSWC04116, oSWC05110, oSWC05111, oSWC05112, and oSWC05113) to confirm successful tagging.

Table 2. 2 All primers used in this study

Names	Sequence (5'-3')	Description
primers for making HrpL-FLAG		
oSWC04110	ATTATCTAGAGGTCGCGCTGAATGAGGATGC	Forward primer for applification of flankA of HrpL
oSWC04112	TCACTTGTCATCGTCGTCCTTGTAGTCGGCGAACGGGTCGATTTGCTG	Reverse primer for applification of flankA of HrpL
oSWC04114	GACTACAAGGACGACGATGACAAGTGAATGGCGATCTCGATCATTTTTT	Forward primer for applification of flankB of HrpL
oSWC04116	ATTATCTAGAGGGTTACTTTGTGCCAACCCCTTCA	Reverse primer for applification of flankB of HrpL
oSWC05110	GCGCAATACCCTGGCCGCGTTATCCA	Forward primer for checking the whole manipulated region of HrpL surrounding flankA
oSWC05111	GGTGAGTTGCCATCCGAGAGTGA	Reverse primer for checking the whole manipulated region of HrpL surrounding flankA
oSWC05112	GGAAGTGTCGCTGGAGATGGA	Forward primer for checking the whole manipulated region of HrpL surrounding flankB
oSWC05113	CTGTGCGCAGCCTTCTTG	Reverse primer for checking the whole manipulated region of HrpL surrounding flankB
Adapters and primers for ChIP-Seq library preparation		
P1 adapter	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	part of sequence pairs of P1 adaptor for ChIP-exo
P1 adapter	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	part of sequence pairs of P1 adaptor for ChIP-exo
P2 primer #1	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	P2 primer for ChIP-exo
P2 adapter #1	GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATA	part of sequence pairs of P2 adaptor for ChIP-exo
P2 adapter #1	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	part of sequence pairs of P2 adaptor for ChIP-exo
P2 adapter #2	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	part of sequence pairs of P2 adaptor for ChIP-exo
Primers for making mutants		
HL109	ATTATCTAGAGCGGCCCAGGTAATTGACGGT	Forward primer for applification of flankA of PSPTO_5633
HL110	CAGCAGACTGGTCTTGTGAAGCCTGT	Reverse primer for applification of flankA of PSPTO_5633
HL111	CTTCACAAGACCAGTCTGCTGCTGCGGCTAAACGAGGACTGATGAG	Forward primer for applification of flankB of PSPTO_5633
HL112	ATTATCTAGACCAAACACGATCCGGCTGAACATT	Reverse primer for applification of flankB of PSPTO_5633
HL113	CACTCGGGTAAGCGGTGCTGTTGATT	Forward primer for checking deletion of PSPTO_5633
HL114	TAGTGGCGCTTGAGGTTGGTATCG	Reverse primer for checking deletion of PSPTO_5633

HL115	CCTGTTGATAGGTCTTGGTGCGGT	Forward primer for checking the whole manipulated region of PSPTO_5633
HL116	TTGGTCTTCAGCGTTCTGGCGTC	Reverse primer for checking the whole manipulated region of PSPTO_5633
HL117	ATTATCTAGACATCATTGCCACTGTCTGTTGCG	Forward primer for applification of flankA of PSPTO_0371
HL118	CCATTCCTTTTCTACATCGTAGGCAGTCA	Reverse primer for applification of flankA of PSPTO_0371
HL119	TACGATGTAGAAAAGGAATGGATCATTGACCGCAGGAATTGAAT	Forward primer for applification of flankB of PSPTO_0371
HL120	ATTATCTAGAGTATTACGCAAGTGCTTCAATCGCAATC	Reverse primer for applification of flankB of PSPTO_0371
HL121	GTTCTGTCGATGGGGCGGTCTTC	Forward primer for checking deletion of PSPTO_0371
HL122	AGGTGGCATGTGCTTTAGCGAACAAG	Reverse primer for checking deletion of PSPTO_0371
HL161	CGTGTCGCCCGCCTGACCTG	Forward primer #2 for checking deletion of PSPTO_0371
HL162	GTTCGGAATTGCCTTGCCAGC	Forward primer #3 for checking deletion of PSPTO_0371
HL123	CATCGCCGTGGCAGTCTCTTCA	Forward primer for checking the whole manipulated region of PSPTO_0371
HL124	TGTCTACACTTTCCGGGTTTCTCTCTAAC	Reverse primer for checking the whole manipulated region of PSPTO_0371
HL125	ATTATCTAGATGCAGACCACCGCCCGTCA	Forward primer for applification of flankA of PSPTO_2130
HL126	CGCCAGAACCAGTCGCACAGT	Reverse primer for applification of flankA of PSPTO_2130
HL127	ACTGTGCGACTGGTTCTGGCGTCGGCGAGCGACTGGCG	Forward primer for applification of flankB of PSPTO_2130
HL128	ATTATCTAGACGCAGTACCCAGGGCAATTTCGGCAA	Reverse primer for applification of flankB of PSPTO_2130
HL129	GCGTGAAGCGGGCATCATC	Forward primer for checking deletion of PSPTO_2130
HL130	CAGCAGGGCGTCCTCGCGCTCG	Reverse primer for checking deletion of PSPTO_2130
HL163	CTTTCGGTACGCTGGTGCTGG	Forward primer #2 for checking deletion of PSPTO_2130
HL164	TGCGTTCCAGGTGGTGACATAGC	Reverse primer #2 for checking deletion of PSPTO_2130
HL131	CGTCTGCTGGTCGATACGGGTCAG	Forward primer for checking the whole manipulated region of PSPTO_2130
HL132	CGTGCGCAGTTCATGGCTCA	Reverse primer for checking the whole manipulated region of PSPTO_2130
HL133	ATTAGGATCCGCAGGCACTGTGGAAAGTCGCTG	Forward primer for applification of flankA of PSPTO_2691
HL134	ACTTGTCGCCAGTTCCAATAGGTATTCC	Reverse primer for applification of flankA of PSPTO_2691
HL135	TACCTATTGGAACTGGCGACAAGTGCTCAGGCATCCACTCACTGATAAACG	Forward primer for applification of flankB of PSPTO_2691
HL136	ATTAGGATCCGGCAGGCTTGGCGACAACATCT	Reverse primer for applification of flankB of PSPTO_2691
HL137	ACGGTTTGCGCTGTCAGGTTCA	Forward primer for checking deletion of PSPTO_2691

HL138	ATGTTCACTCACTTCAACTGATGCCG	Reverse primer for checking deletion of PSPTO_2691
HL159	CCTGCTTTGGACAATGCGTCAG	Reverse primer for checking deletion of PSPTO_26912
HL139	CTGCCTCGAACCACGGTCTCCTG	Forward primer for checking the whole manipulated region of PSPTO_2691
HL140	CGGCCAACATCGAAGCCATTATTC	Reverse primer for checking the whole manipulated region of PSPTO_2691
HL141	ATTATCTAGACGGGCATCACGACGTACTCTCCA	Forward primer for applification of flankA of PSPTO_2696
HL142	TTTGCAGATGACTGTCGCTCGTTG	Reverse primer for applification of flankA of PSPTO_2696
HL143	ACGAGCGACAGTCATCTGCAAAGGTTTACTCAGCGCTAACTAGAACAGCC	Forward primer for applification of flankB of PSPTO_2696
HL144	ATTATCTAGACTCAGGGAATCAGCCGATCACG	Reverse primer for applification of flankB of PSPTO_2696
HL145	AGCGGATGCTCAAGGGCGAC	Forward primer for checking deletion of PSPTO_2696
HL146	CTCGAACGGGCCGAGGTAATGCTT	Reverse primer for checking deletion of PSPTO_2696
HL165	TGAGTTGATGCTCCTGGGCGAT	Reverse primer #2 for checking deletion of PSPTO_2696
HL147	TGTACAGAGGGCGACGAGCAGTATT	Forward primer for checking the whole manipulated region of PSPTO_2696
HL148	TCCAATCGCGCCTACCAAACAAATC	Reverse primer for checking the whole manipulated region of PSPTO_2696
HL149	ATTATCTAGAAGTCCTGGTACCTGATCAGCGATTCTTA	Forward primer for applification of flankA of PSPTO_3331
HL150	CACGATTCGAACAAAGTAATTGATATTCA	Reverse primer for applification of flankA of PSPTO_3331
HL151	CAATTACTTTGTTCGAATCGTGCTTGGACGGTTCGCCGATTAAGC	Forward primer for applification of flankB of PSPTO_3331
HL152	ATTATCTAGAGCGGCAGGCGCAGGATCAG	Reverse primer for applification of flankB of PSPTO_3331
HL153	CTGCCGCCAACGACCTCAAAG	Forward primer for checking deletion of PSPTO_3331
HL154	GGATGAAGGACAGGTCGCCCAGC	Reverse primer for checking deletion of PSPTO_3331
HL155	CGGGCATATGACCTTTGGTAACTACAGT	Forward primer for checking the whole manipulated region of PSPTO_3331
HL156	GCCGTACAAGGCGCGGGC	Reverse primer for checking the whole manipulated region of PSPTO_3331
HL160	GCGTAGGTCTTGAGCAAATGGTC	Reverse primer #2 for checking deletion of PSPTO_3331
HL397	ATTATCTAGAGACCGTGACCAAAGGCCCGA	Forward primer for applification of flankA of PSPTO_5240
HL398	*CGAGGGCTGCAAGGTTA*CACGCATCAGTC	Reverse primer for applification of flankA of PSPTO_5240
HL399	TAACCTTGCAGCCCTCGGTCAGTGCCGGTATGGATGCG	Forward primer for applification of flankB of PSPTO_5240
HL400	ATTATCTAGACATGGGCTTACTTTCTGTACGAGTGGG	Reverse primer for applification of flankB of PSPTO_5240
HL401	TCACCGATGACGACATCAATGCG	Forward primer for checking deletion of PSPTO_5240

HL402	GACCCGTAAGCGAACAGCCAGAC	Reverse primer for checking deletion of PSPTO_5240
HL403	TGATCAAGTGCCGTGGCAGCAA	Forward primer #2 for checking deletion of PSPTO_5240
HL404	ATCGGCGCGTGCATCCTG	Reverse primer #2 for checking deletion of PSPTO_5240
HL405	AGGAGATCGTCATCGAGGGTGAT	Forward primer for checking the whole manipulated region of PSPTO_5240
HL406	TACGGCGTTCTGGACTGGGC	Reverse primer for checking the whole manipulated region of PSPTO_5240
HL419	ATTATCTAGATCCGACACCAACAATAAATTCAACAATCT	Forward primer for applification of flankA of PSPTO_0816
HL420	TAACGTAAAGCCCCTGGATGTCGC	Reverse primer for applification of flankA of PSPTO_0816
HL421	TCCAGGGGCTTTACGTTATGCTGGGGCCGCTAAAACCG	Forward primer for applification of flankB of PSPTO_0816
HL422	ATTATCTAGATTCGATGGCCGACGCCTTGAGACTTTC	Reverse primer for applification of flankB of PSPTO_0816
HL423	ATCAAGTTATTCGAGGGTGTCAGCAAC	Forward primer for checking the whole manipulated region of PSPTO_0816
HL424	CTTCGGGCGAGCCCGGACGC	Reverse primer for checking the whole manipulated region of PSPTO_0816
HL425	CAGGATTTCTACGCGATATGGGATG	Forward primer for checking deletion of PSPTO_0816
HL426	TCACTGCCGAGATATCGACACTGC	Reverse primer for checking deletion of PSPTO_0816
HL447	ATTAGAATTCTGTCGGTCTGGCTCAGTGTTTCCTG	Forward primer for applification of flankA of PSPTO_5633 on plasmid B
HL448	CAGCAGACTGGTCTTGTGAAGCCTG	Reverse primer for applification of flankA of PSPTO_5633 on plasmid B
HL449	CACAAGACCAGTCTGCTGCTGCGGCTAAACGAGGACTGATGAGTT	Forward primer for applification of flankB of PSPTO_5633 on plasmid B
HL450	ATTAGAATTCAGTGGGACAACATGGTGATTGTCGTG	Reverse primer for applification of flankB of PSPTO_5633 on plasmid B
HL451	GCGGCCACAGGGTGTCAGG	Forward primer for checking deletion of PSPTO_5633 on plasmid B
HL452	TGGATGCAGGTTCGTTTGCAGTCA	Reverse primer for checking deletion of PSPTO_5633 on plasmid B
HL453	GACGCGGTTGGCCAGAAATCC	Forward primer for checking the whole manipulated region of PSPTO_5633 on plasmid B
HL454	CGACGTGGCCTCGGGTGAC	Reverse primer for checking the whole manipulated region of PSPTO_5633 on plasmid B
HL427	ATTATCTAGAAGGACATCGTCATGGCGTGTGAT	Forward primer for applification of flankA of PSPTO_0871
HL428	CGGGGAAAGAGTCTGACCGGAG	Reverse primer for applification of flankA of PSPTO_0871
HL429	GGTCAGACTCTTTCCCCGGCCGTGCGCCGTCTTGAAG	Forward primer for applification of flankB of PSPTO_0871
HL430	ATTATCTAGAGACACGAATATCATCCATGATTCCGA	Reverse primer for applification of flankB of PSPTO_0871

HL431	TTCCGGTCCACAGCGTGTTGC	Forward primer for checking the whole manipulated region of PSPTO_0871
HL432	CCTCTGGGAAGCGTCGCCCGGA	Reverse primer for checking the whole manipulated region of PSPTO_0871
HL433	ATGTCCACAGCATGACCAAGGGC	Forward primer for checking deletion of PSPTO_0871
HL434	GGAAGGTCTGAAGCAGGCATCAAC	Reverse primer for checking deletion of PSPTO_0871
HL435	TCAAAGAGTGTCATAGGATACCAACTGC	Forward primer #2 for checking deletion of PSPTO_0871
HL436	AGCGGCTCGGACGTACAACAATC	Reverse primer #2 for checking deletion of PSPTO_0871
Promoter fusion primers		
HL208	CACCCGATGTTCATTGCTCCCCG	forward primer to amplify promoter region of PSPTO_5633
HL209	ACGATTGCTTGTAGAGAGCACGGAAG	reverse primer to amplify promoter region of PSPTO_5633
HL210	CACCGCGGATAGGCAACAAGCTCGAC	forward primer to amplify promoter region of PSPTO_0371
HL211	TGCCCACCCGTCAGCAGCG	reverse primer to amplify promoter region of PSPTO_0371
HL212	CACCTGCCGCCCGTTGCTGCGTGGC	forward primer to amplify promoter region of PSPTO_1645
HL213	GTTGGCTGGCCGGGCTGCCG	reverse primer to amplify promoter region of PSPTO_1645
HL216	CACCCATCAGGCCGCGACCAAAG	forward primer to amplify promoter region of PSPTO_2691
HL217	CTTTATCTGAATTTTTTTCGGTTTCAGC	reverse primer to amplify promoter region of PSPTO_2691
HL222	CACCCATGCCGTACGTCGCAATACG	forward primer to amplify promoter region of PSPTO_2696
HL223	TATTTGACGAGGTCAACCATGCAGC	reverse primer to amplify promoter region of PSPTO_2696
HL224	CACCGCTGGGGACTTTCTGGTGGGA	forward primer to amplify promoter region of PSPTO_3331
HL225	GCACGATTCGAACAAAGTAATTGATATTC	reverse primer to amplify promoter region of PSPTO_3331
HL228	CACCAGCATTACCTTTTCAGGGATGATTTAGA	forward primer to amplify promoter region of PSPTO_3948_3949
HL229	ATCATGCGTTAGGTATTCCATCTCATC	reverse primer to amplify promoter region of PSPTO_3948_3949
HL230	CACCACTTAAAAATAAAGTTCATGGCATGGAAC	forward primer to amplify promoter region of PSPTO_4699
HL231	CGGGGAATATCATCTGCGTCCTTG	reverse primer to amplify promoter region of PSPTO_4699
HL232	CACCTGAAATTTCTATGTACCGAGGGAACG	forward primer to amplify promoter region of PSPTO_4721
HL233	TGTGACAGGCTTGGCATTTTTTG	reverse primer to amplify promoter region of PSPTO_4721
HL236	CACCTCATTGAACCGGCAGCGCCT	forward primer to amplify promoter region of PSPTO_4955

HL237	GATGAATGTCCTGAAAATTTTTGAGAAC	reverse primer to amplify promoter region of PSPTO_4955
HL238	CACCTGGTATCAACTGACGCTGGAACCCA	forward primer to amplify promoter region of PSPTO_5053
HL239	CGCATAAGCCGACACTTCGTACTG	reverse primer to amplify promoter region of PSPTO_5053
HL268	CACCGCCAATGCCAGTTGCGTAGAAC	forward primer to amplify promoter region of PSPTO_2130
HL269	GTCAGGCGGCCACCCAGC	reverse primer to amplify promoter region of PSPTO_2130
HL270	CACCCAGCGCATCCTGCCGAGTG	forward primer to amplify promoter region of PSPTO_4750 (antisense)
HL271	CAGCAGTACGCAAGTGCCTGGATGAAGAG	reverse primer to amplify promoter region of PSPTO_4750 (antisense)
HL327	CACCTGGCTCATTCAGGCGCTAACAAC	forward primer to amplify promoter region of PSPTO_0871
HL328	CATCCAGAAACGCACCTGCCAA	reverse primer to amplify promoter region of PSPTO_0871
HL329	CACCCAGGCGGTCAATGGCAGTGTT	forward primer to amplify promoter region of PSPTO_0816
HL330	TGGCAGCATTGAGTTCAAATACCC	reverse primer to amplify promoter region of PSPTO_0816
HL381	CACCTGTTTTCTTGATCCAGATCAATAAAGAC	forward primer to amplify promoter region of hopAT1
HL382	ATTATTAGCCCTTTTAATTGATCTTCTCCT	reverse primer to amplify promoter region of hopAT1
HL387	CACCGCGTCCGGCAGCCTGAGCG	forward primer to amplify promoter region of PSPTO_PSPTO_1843
HL388	GTGGCTGAAGTCAAAGCGTAGACGC	reverse primer to amplify promoter region of PSPTO_PSPTO_1843
HL389	CACCGGTAATAGCGAGCGCGGCTTCCAGA	forward primer to amplify promoter region of PSPTO_PSPTO_3481
HL390	CGTTCAGCGTCGATCCTGTTTGC	reverse primer to amplify promoter region of PSPTO_PSPTO_3481
HL391	CACCTCGGTCGAAAAGCAGATTAACGAGGT	forward primer to amplify promoter region of PSPTO_3721
HL392	AGCGCCATGGCAATCATCACG	reverse primer to amplify promoter region of PSPTO_3721
HL393	ATGACCGAGCAACCCTTCTCCCTTC	forward primer to amplify promoter region of PSPTO_4340
HL394	GCGTTGATGGCTTTACTTAACACCTGAT	reverse primer to amplify promoter region of PSPTO_4340
HL437	CACCGGCCTGCTCCGCTCATCAG	forward primer to amplify promoter region of PSPTO_4340
HL438	GGCACAGTGCCTGAGGACAAGGGCA	reverse primer to amplify promoter region of PSPTO_4340
HL442	CACCGTCATGATCGACAACACGCCCAT	forward primer to amplify promoter region of PSPTO_5240
HL443	GGTTACACGCATCAGTCTATTCCCAAC	reverse primer to amplify promoter region of PSPTO_5240
RT-PCR primers downstream of hrp promoters		
HL051	GTTGTCCTGTCCCGCAAAGC	forward primer for quantitative RT_PCR of transcript from hrp promoter

		upstream of PSPTO_5633
HL052	GGAGTGTTTCTTGTGTGGTTGAG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_5633
HL053	GTACCGATCATGGACAAGGAAGG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0371
HL054	CTGGTGCCACTGGTAAAGCC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0371
HL055	TGCCGCCTTATGTCATTTTCCC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_1644
HL056	CACCCACGCCACTCACTTTG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_1644
HL057	CGGTGGGCTATCTGCTCAAGG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2130
HL058	CGCTGTTCAATGCCTCTTCGC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2130
HL059	GTTCAACCAGGCGTTCTCGTG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2691
HL060	CGGTCTTGATGTCCATGCTGTG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2691
HL061	ATGAAACAACGAGCGACAG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2696
HL062	GCTTCAATCTTGCCACCTG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_2696
HL063	AATCGTGCCAGTTGCCGTTG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3331
HL064	GCCATTGACCGCTGAGTTCG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3331
HL065	GATGGAATACCTAACGCATGATGG	forward primer for quantitative RT_PCR of transcript from hrp promoter between PSPTO_3948 and PSPTO_3949
HL066	CCCTTCAGAAAGTCAGGACACAG	reverse primer for quantitative RT_PCR of transcript from hrp promoter between PSPTO_3948 and PSPTO_3949
HL071	GCATGAGACTGGAGATCAATGGTG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4699
HL072	GGAACAAGGCAAGCCGTAGTG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4699
HL073	GCTATCCAGTGGCAGGAAGGC	forward primer for quantitative RT_PCR

		of transcript from hrp promoter upstream of PSPTO_4721
HL074	TCTTTAGGGTCTGTACGCAGTTCTATG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4721
HL075	GCGTACTGCTGCCCGAAG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4750
HL076	GCCGCCATTACACAAGGACAG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4750
HL077	CCTGTCGCCGAAGGACTACC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4955
HL078	GTCTGGTTGACCGAGAAGATACG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4955
HL079	GTTGCCGCTGGACGAATTGC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_5053
HL080	CGAGGCTTAGACCCGTTCTTTCC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_5053
HL333	CGCCAGTTTGCTTTCCGAACAGG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0816
HL334	GCATTGCCGCCAACCACACC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0816
HL339	TTTACTGCGGTGCTTCCTC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of hopAT1
HL340	TGTAATAAAGTAGCGATGGTTGTG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of hopAT1
HL349	TGCCTGCGGTTTCCTACAC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_1843
HL350	AGGTCGGAACGAATCTTCTGG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_1843
HL355	CGTTGTGCTTCTTCAATTCGTC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3481
HL356	TCAGGTCCATCGTGTAACTCC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3481
HL361	CTGCTGATTGCCGACGAACC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3721
HL362	CAACAGTGCCATGCCCAGTC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_3721

HL371	CACTGGCGTGAAGGTGCTTTGTTG	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4340
HL372	GGAAGGCAGCGTGGCGGAACC	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_4340
HL375	GCTCAACTGGCAGGCAAAAC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_5240
HL376	CCACGGCGAGATTCATACG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_5240
HL377	GCTACCGTATTCCCCAGGCATCC	forward primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0871
HL378	CGCAAGCCAGGGTGACTCCAG	reverse primer for quantitative RT_PCR of transcript from hrp promoter upstream of PSPTO_0871
oSWC05280	GCCGATCAGATTCAGATGCTCAG	forward primer for quantitative RT_PCR of PSPTO_1404 (hrpL)
oSWC05281	TTACGCAGGGCTTCAAGAAACAC	reverse primer for quantitative RT_PCR of PSPTO_1404 (hrpL)
oSWC06207A	GCGAATGGAAGCGAGTGTCTG	Forward primer for quantitative RT_PCR for gene shcN_1369
oSWC06208A	CGTTGTCCTGCGGCATCG	Reverse primer for quantitative RT_PCR for gene shcN_1369
oSWC06209A	CTGCCAATTTCAAAATCAGCGAAG	Forward primer for quantitative RT_PCR for gene shcM_1374
oSWC06210A	CCCATGTCACCGTAGATCAGAAC	Reverse primer for quantitative RT_PCR for gene shcM_1374
oSWC06211A	AGGAAAGAGCCGAACAACCC	Forward primer for quantitative RT_PCR for gene avrE1_1377
oSWC06212A	TTGGATTTCTGGAACAATTTACCG	Reverse primer for quantitative RT_PCR for gene avrE1_1377
oSWC06213A	GCACCGACTGAAGCGACTG	Forward primer for quantitative RT_PCR for gene hrpK1_1405
oSWC06214A	TGCGTCTGCGGCTTTGGG	Reverse primer for quantitative RT_PCR for gene hrpK1_1405
oSWC06217A	ACAGCCTGTCGGTTGATGC	Forward primer for quantitative RT_PCR for gene hrpA1_1381
oSWC06218A	GGTCTGCTTCTTCATTGTTTCCTG	Reverse primer for quantitative RT_PCR for gene hrpA1_1381
oSWC06219A	AGCGTCAGACATCAACTTGCG	Forward primer for quantitative RT_PCR for gene avrPto1_4001
oSWC06220A	AGATTGTTGTAGCGATTTCTCAGG	Reverse primer for quantitative RT_PCR for gene avrPto1_4001
oSWC06221A	GCAAAGGAAAATCTATCGCACCAAC	Forward primer for quantitative RT_PCR for gene hopE1_4331
oSWC06222A	CGGCACTTCTGTACGCAATCTG	Reverse primer for quantitative RT_PCR for gene hopE1_4331
oSWC06223A	CCTGCTGTTGATCGGCTTGC	Forward primer for quantitative RT_PCR for gene shcA_5353
oSWC06224A	CATTGAGGAGCGGATTGAGAGC	Reverse primer for quantitative RT_PCR for gene shcA_5353

oSWC06225A	CGTACTGTTCATAACTTCTATCGGACTG	Forward primer for quantitative RT_PCR for gene hopC1_0589
oSWC06226A	ATAAACTTTCTGTGCCAGCCATCG	Reverse primer for quantitative RT_PCR for gene hopC1_0589
oSWC06231A	CTTTACCAAAGGTGAAGGAAGCAAC	Forward primer for quantitative RT_PCR for gene hopH1_0588
oSWC06232A	ACGGTGAGCCATTCTGATTTAAGC	Reverse primer for quantitative RT_PCR for gene hopH1_0588
oSWC06237A	CCAGTTTGAACAGTCAGGTGCTC	Forward primer for quantitative RT_PCR for gene hopR1_0883
oSWC06238A	TTAATGACACTCTCGCCATCGG	Reverse primer for quantitative RT_PCR for gene hopR1_0883
oSWC06239A	CCATTGGAGGGGCATGAAGAGG	Forward primer for quantitative RT_PCR for gene hopN1_1370
oSWC06240A	GTTGTACCTGCTCGGGACTGG	Reverse primer for quantitative RT_PCR for gene hopN1_1370
oSWC06241A	CAACGCAGGCTTTGGCAACC	Forward primer for quantitative RT_PCR for gene hopAA1-1_1372
oSWC06242A	CTTGGGCTCTTTATCCTTCAAACCG	Reverse primer for quantitative RT_PCR for gene hopAA1-1_1372
oSWC06243A	GCAGGACGCCAGCAAGCC	Forward primer for quantitative RT_PCR for gene hrpW1_1373
oSWC06244A	TTTATTGGAGTTGGTGAGCATCTG	Reverse primer for quantitative RT_PCR for gene hrpW1_1373
oSWC06245A	ACCAGGCAGCGATGGAAGC	Forward primer for quantitative RT_PCR for gene hrpF_1387
oSWC06246A	GCACGCAGGCATTCATTGACG	Reverse primer for quantitative RT_PCR for gene hrpF_1387
oSWC06249A	GAGTGGCGGGCATAACGG	Forward primer for quantitative RT_PCR for gene hopP1_2678
oSWC06250A	GGCGGAGCGAAATTGACCTG	Reverse primer for quantitative RT_PCR for gene hopP1_2678
oSWC06251A	CCGACTGACACCAGCACATC	Forward primer for quantitative RT_PCR for gene hopAK1_4101
oSWC06252A	GAGCAGGTTGTTGAGGATATTACG	Reverse primer for quantitative RT_PCR for gene hopAK1_4101
oSWC06253A	GTTACACGCACGACCTCACTG	Forward primer for quantitative RT_PCR for gene shcS2_4589
oSWC06254A	CGATGACCTGCTTGATCTGCTC	Reverse primer for quantitative RT_PCR for gene shcS2_4589
oSWC06255A	CTGTCCCCTGTCCATACGATGAG	Forward primer for quantitative RT_PCR for gene shcS1_4599
oSWC06256A	CCCGCTCCCGCCAAGGTC	Reverse primer for quantitative RT_PCR for gene shcS1_4599
oSWC06257A	ACAGGTTTTCGGTAATAGGCTCAC	Forward primer for quantitative RT_PCR for gene hopAA1-2_4718
oSWC06258A	TTCGGAGTACAGGCGGTTGG	Reverse primer for quantitative RT_PCR for gene hopAA1-2_4718
oSWC06259A	CTCAGCCTAATACCGATCACCAAC	Forward primer for quantitative RT_PCR for gene hopAO1_4722
oSWC06260A	GGTTACTTGCGTGTCGTTACTGG	Reverse primer for quantitative RT_PCR for gene hopAO1_4722
oSWC06263A	TGATGGGGAGCCCGTCTGAG	Forward primer for quantitative RT_PCR for gene hopK1_0044

oSWC06264A	TGTTGGAAGAGCCTGCTTGAGTG	Reverse primer for quantitative RT_PCR for gene hopK1_0044
oSWC06265A	AAGGGCTCGTCCGCACAAG	Forward primer for quantitative RT_PCR for gene hopY1_0061
oSWC06266A	CGCTACGCACCTTATTCAACCAC	Reverse primer for quantitative RT_PCR for gene hopY1_0061
oSWC06273A	CGGGCACGAAGCAGTTGG	Forward primer for quantitative RT_PCR for gene PSPTO_0834
oSWC06274A	GTTCGCAAACGCCACAGTATTC	Reverse primer for quantitative RT_PCR for gene PSPTO_0834
oSWC06279A	GCCCAGCAATCCCATCCTAAAAG	Forward primer for quantitative RT_PCR for gene hopD1_0876
oSWC06280A	GCGTTGGTGTTCGGCATTAGC	Reverse primer for quantitative RT_PCR for gene hopD1_0876
oSWC06281A	AGCATGAGCAGCCCAGTTTTG	Forward primer for quantitative RT_PCR for gene hopQ1-1_0877
oSWC06282A	GAGGTCAGATGAGCGTTACAGG	Reverse primer for quantitative RT_PCR for gene hopQ1-1_0877
oSWC06285A	CAAGTTAAATCGGCAGGAACCTC	Forward primer for quantitative RT_PCR for gene hopAM1-1_1022
oSWC06286A	GCCTAAATCACTGAGTCTGTCTG	Reverse primer for quantitative RT_PCR for gene hopAM1-1_1022
oSWC06287A	GTGGGCTGCCAGTCTGTG	Forward primer for quantitative RT_PCR for gene hrpH_1378
oSWC06288A	GCGAACCTTCCTTATCATCTCC	Reverse primer for quantitative RT_PCR for gene hrpH_1378
oSWC06289A	CGTTCGGTACTGAGCAAGTCG	Forward primer for quantitative RT_PCR for gene hrpO_1399
oSWC06290A	TGACCAGGTTCGGCACTTACG	Reverse primer for quantitative RT_PCR for gene hrpO_1399
oSWC06291A	GCCACTTCCCAATGCTGATCTG	Forward primer for quantitative RT_PCR for gene hrpJ_1403
oSWC06292A	CTGAATGAGGATGCGTCTGTGC	Reverse primer for quantitative RT_PCR for gene hrpJ_1403
oSWC06301A	ATGATCCTAACTATGCTGCTCGTG	Forward primer for quantitative RT_PCR for gene hopAD1_4691
oSWC06302A	ACCTGTGCTTTCAAATGTTCTGC	Reverse primer for quantitative RT_PCR for gene hopAD1_4691
oSWC06303A	CCACAACATCTATCACCCAAACCC	Forward primer for quantitative RT_PCR for gene hopAQ1_4703
oSWC06304A	CCGACCGAGATTAAATGCGACTG	Reverse primer for quantitative RT_PCR for gene hopAQ1_4703
oSWC06305A	CGCTCACCAGGCTCTAAAACG	Forward primer for quantitative RT_PCR for gene hopD_4724
oSWC06306A	GATTGAAATTGAAGTGTGCCATCC	Reverse primer for quantitative RT_PCR for gene hopD_4724
oSWC06307A	GCATGATGGCAAAGGTGGTTTTAC	Forward primer for quantitative RT_PCR for gene hopG1_4727
oSWC06308A	TTTCAAGAGTGCTGCTATGGTTG	Reverse primer for quantitative RT_PCR for gene hopG1_4727
oSWC06311A	CCACATCGCTGGCAAAGTCG	Forward primer for quantitative RT_PCR for gene hopI1_4776
oSWC06312A	CGAGAGCCCTTTCTGCTGACG	Reverse primer for quantitative RT_PCR for gene hopI1_4776

Binding site RT-PCR primers		
HL083	TGGTGCGGCTCCACTAAC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_5633
HL084	AGAATGAATGACTAGGCTCTATACG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_5633
HL085	AAGCGGATAGGCAACAAGC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0371
HL086	AGCGAAGGGCGGACAAAG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0371
HL087	CGACCTTGAAGGCTACGG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_1644
HL088	TTGAAGTTGTTTGTGGTTTGAG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_1644
HL089	AATCGCACTGTTCCGCATC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2130
HL090	GTTTCCGTCACTCAGTTCCAG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2130
HL091	CAGCCAACACTGCCAAAC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2691
HL092	ACCGAAAGTTCCTGTCTGG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2691
HL093	CGGGAACTGTTTTCTGGCAATGG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2696
HL094	GCAGCGTGAAAGCGAGAGATTG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_2696
HL095	TGAACTACAACCAGAGCAGCAACC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3331
HL096	GTGGCAAACGCTTGACCAACAG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3331
HL097	TCAGGGATGATTTAGAGGAAG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3947
HL098	CCTAACGCTTGGATAAGTTC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3947
HL099	CATGGCATGGAACCGATAAC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4699
HL100	ATCTGCGTCCTTGGTGTG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4699
HL101	TCCTTTCATAGCGTCAGC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4721
HL102	ATAGCAATATCTGTACTAAGTGG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4721
HL103	GTGATCGCTTGCTGGAACC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4750
HL104	CAAGTGCCTGGATGAAGAGC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4750
HL105	AACCATTGAGCCTGTTCGC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4955
HL106	GGTCCTGAAGATGAGTGTCCTG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4955
HL107	CGCACTTGTCCTGGTATCAACTG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_5053
HL108	TCCTGAATGTCCCAAAGAATGTCG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_5053

HL335	GGCGGTCAATGGCAGTGTTGTCAG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0816
HL336	CGGCGGAAAGCGGCATGTACC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0816
HL337	AATTGCTGGCTACGCTAACC	forward RT-PCRprimer at ChIP-Seq peak upstream of hopAT1
HL338	GTAAACCCACCATCTACAAATCG	reverse RT-PCRprimer at ChIP-Seq peak upstream of hopAT1
HL345	GCGTACAGGTTGAAACACAG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_1843
HL346	CTCGCCCAGCACTTGACG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_1843
HL353	CCAGGATCGGTGACTAACAG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3481
HL354	GCGACGAATTGAAGAAGCAC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3481
HL359	AGCCACATAAACGCCTCAAAGC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3721
HL360	CGGTTCGTCGGCAATCAGCAG	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_3721
HL367	AAGGCTGATCTGTTGAAAGTGAAAGTTG	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4340
HL368	CTGAGGACAAGGGCAAGACTATTTACC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_4340
HL379	ATCATTATTGACGGGCATAGC	forward RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0871
HL380	CTTGTGCTGATGGGTTCC	reverse RT-PCRprimer at ChIP-Seq peak upstream of PSPTO_0871
primers for other purposes		
oSWC750	ACTAAGCTTATGGTCCGTCCTGTAGAAACCCC AAC	PCR 5' *iucD*; includes *HindIII* site.
oSWC751	AGTAAGCTTATTATTGTTTGCCTCCCTGCTGC	PCR 3' iucD; includes *HindIII* site.
oSWC463	CACCTAGGGCAGTTCTAAGCCGGATTATG	PCR upstream *hrpJ ;* includes CACC for directional cloning in pENTR/D.
oSWC464	AGGATCACAATCTTCTGAAACATG	PCR upstream *hrpJ.*
oSWC47	GGAATTCTGATTGATTGAGAAGGAGATATAC ATATGAGTAAAGGAGAAGAACTTTTC	PCR 5' *gfp* adds *EcoRI* site, 3 stop codons and *SD*.
oSWC48	GGAATTCCTATTTGTATAGTTCATCCATGCC	PCR 3' *gfp* adds *EcoRI* site.
oSWC00381	CCGCAAGGTGATTATCTCAGC	Forward primer for RT-PCR of *gap1*
oSWC00382	TGGAGATGATCTGGTGCGACT	Reverse primer for RT-PCR of *gap1*
oSWC00379	AGGCAAGTATTTCCTGTCGCC	Forward primer for RT-PCR of *gyrA*
oSWC00380	CTGGTACTCACCCAGCAGTTTTT	Reverse primer for RT-PCR of *gyrA*

Mutants constructed for this study. A uniform strategy was used to construct deletions in *PSPTO_5633, PSPTO_0371, PSPTO_2130, PSPTO_2691, PSPTO_2696, PSPTO_3331,* and *PSPTO_5240.* Regions flanking the gene of interest, designated flank A and flank B, were amplified from DC3000 genomic DNA by two primer pairs (see Table 2.2), purified by gel electrophoresis (Qiagen), and joined by SOEing PCR [147]. The joined fragment was then digested with XbaI (or BamHI for *PSPTO_2691*), and ligated with XbaI (or BamHI) digested pK18mobsacB. The resulting constructs were sequenced to confirm correct structure using primers M13F, M13R, and the two primers flanking the deletion region. The deletion constructs were introduced into DC3000, and *ΔhopQ1-1* by electroporation. Merodiploid intermediates were selected for growth on medium containing kanamycin. Recombinants that had eliminated pK18mobsacB plasmid sequences were identified by sucrose counter-selection. Sucrose-resistant, Kan-sensitive colonies were screened by PCR using two primers flanking the deletion region. Mutants were confirmed by Sanger sequencing using four to six primers covering the manipulated region (see Table 2.1 for all mutants).

Construction of P_{hrpJ}::iucD reporter plasmid. The plasmid pBS181encodes a transcriptional fusion of the *hrpJ* promoter with the *iucD* reporter gene (P_{hrpJ}::*iucD*). The pBS181 plasmid was constructed by Gateway LR recombination (Invitrogen) between the pBS62 entry vector and the pBS180 destination vector. The pBS180 destination vector was constructed by replacing the *lux* operon of pBS59 [63] with the *iucD* gene. The *iucD* open reading frame was PCR amplified from pENTR-Gus (Invitrogen) using oSWC750 and oSWC751, digested with HindIII and ligated with the 8.6 kb fragment of HindIII digested pBS59, generating pBS180. The *hrpJ* promoter region was PCR amplified from DC3000

genomic DNA using oSWC463 and oSWC464 and cloned in pENTR/D-topo (Invitrogen) to yield pBS62. The *hrpJ* promoter *iucD* fusion was then constructed by LR reactions between pBS62 and pBS180 to yield pBS181.

Construction of promoter fusions. The *gfpmut2* gene was amplified from pTGS [140] using oSWC47 and oSWC48 for PCR. These primers introduced EcoRI recognition sequences to the 5' and 3' ends of the PCR product. The 5' primer (oSWC47) also introduced a stop codon (TGA) in each frame (three total) followed by a Shine Delgarno sequence, which precedes the *gfpmut2* start codon by 8 base pairs (bps). The *gfpmut2* PCR product was digested with EcoRI and ligated to similarly digested pUCP24, yielding pBS12. pHL1 was constructed to function as a destination vector compatible with Gateway cloning. The Gateway cassette from plasmid pBS46 [63] was digested with enzyme KpnI and was ligated to plasmid pBS12 digested using the same enzyme. The resulting pHL1 construct contains a Gateway cassette, the promoter-less reporter gene *gfpmut2*, and antibiotic resistance genes for tetracycline and kanamycin resistance. To test for the presence of *hrp* promoters, fragments of DNA (150 bps to 200 bps) containing candidate promoters were cloned from DC3000 genomic DNA using the Expand High Fidelity PCR system (Roche, Basal, Switzerland). The forward primers have four additional bases (CACC) for compatibility with TOPO cloning vectors. PCR fragments were cloned into pENTR/SD/D by directional TOPO cloning and were subsequently used to generate the *gfp* reporter constructs by LR reaction with the pHL1 destination vector (Table 2.1) using LR Clonase II enzyme mix (Invitrogen). All constructs were confirmed by sequencing. Plasmids were transformed by electroporation [145] into DC3000 Δ*pvsA,* which differs from WT DC3000

in that it cannot produce pyoverdine, a fluorescent siderophore that makes accurate measurement of GFP concentration difficult [142].

Construction of PSPTO_5633–Cya fusion. The T3SS dependent translocation reporter gene *adenylate cyclase* [139] was fused to the C terminus of PSPTO_5633 using Gateway cloning (PCPP6413). The plasmids used to generate pCPP6413, containing the PSPTO5633-Cya fusion, are described in Table 2.1. Subsequently pCPP6413 and the translocation reporter control pCPP5388 (AvrPto-Cya) were conjugated into different *Pto* DC3000 strains by tri-parental mating using the helper plasmid pRK2013.

Western blot

Proteins were resolved in a precast 4-20% polyacrylamide gel (Bio-Rad) and transferred to a PVDF membrane (Millipore) by electrophoresis. The membrane was then blocked in 5% non-fat milk for 2 hours at room temperature, and incubated with anti-FLAG M2 antibody (Sigma-Aldrich) for 1 hour at room temperature with gentle shaking. The membrane was washed three times in TBST buffer (50 mM Tris, 150 mM NaCl, 0.05% Tween-20, pH 7.6) and incubated with Alkaline Phosphatase conjugated 2^0Antibody (Millipore) for 1 hour at room temperature. Proteins were detected after adding BCIP/NBT substrate (Sigma Aldrich) at room temperature.

Chromatin immunoprecipitation with exonuclease treatment (ChIP-exo) paired with high-throughput sequencing (ChIP-Seq)

ChIP-exo and ChIP-Seq were performed as described [142, 148]. Briefly, bacterial cultures were harvested and cross-linked with 1% formaldehyde final concentration. After

20 minutes incubation at room temperature with slow shaking, glycine was added at 0.36 M final concentration to quench the cross linking reaction. Cells were collected by centrifugation at 4^0C for 5 minutes at 5,000 x *g* and washed twice in ice cold TBS. Washed pellets were stored at –70^0C until processed. To lyse cells, 1 ml of CelLytic B (Sigma-Aldrich), supplemented with 10 μl LongLife™ Lysozyme (1,500 U, G-Biosciences) and 10 μl PMSF (0.1 M, Sigma-Aldrich), was thoroughly mixed with thawed pellets by vortexing. Cells were incubated at 37^0C for 10-15 minutes, and then disrupted by sonication (6 repetitions, 30 seconds each, with 2 minutes cooling between each pulse). Continuous pulse power at 15% power was used to produce fragments of size around 300 bps. For each ChIP-Seq sample, 40 μl of ANTI-FLAG M2 Affinity Gel (Sigma-Aldrich) was pre-washed in cold TBS, added to the bacterial lysates, and incubated with gentle shaking at 4^0C for 2 hours. Unbound DNA fragments were eluted using two washes with TBS. Note that *hrpL-FLAG* samples were prepared following the ChIP-exo protocol that includes specialized enzyme treatment and library preparation procedures (all steps described below), while the *ΔhrpL* samples were eluted, reverse cross-linked (below) and directly submitted for high-throughput sequencing without additional manipulation as detailed previously [149].

For ChIP-exo, resin-bound DNA from the *hrpL-FLAG* samples were treated using methods adopted from Rhee *et al.* [132], with two TBS washes following each step:

1. End polishing was accomplished using 4.5 U of T4 DNA polymerase, 200 μM dNTPs and 50 μg/ml BSA at 12^0C for 30 minutes. This reaction generates blunt end DNA fragments.

2. P2 adaptor (manufactured by IDT; see Table 2.2 for sequence) was ligated to both ends of the sheared DNA fragments using 1600 U of T4 DNA ligase and 200 ng adaptor at

25^{0}C for 60 minutes. The P2 adaptor has only one blunt end that is available for ligation in this reaction.

3. Nick repair was accomplished using 20 U of Φ29 polymerase with 10 mM dNTPs and BSA at a final concentration of 100 ug/ml at 30^{0}C for 20 minutes. This step generates double stranded DNA fragments without nicks.

4. The sample was digested with 10 U of λ exonuclease at 37^{0}C for 30 minutes. This enzyme preferentially degrades double-stranded DNA from the 5′-end but is unable to degrade DNA regions protected by cross-linked protein. Unprotected double-stranded DNA is degraded in this step, yielding single-stranded DNA.

5. The sample was digested using 30 U of $RecJ_f$ exonuclease at 37^{0}C for 30 minutes. This enzyme is a single-stranded DNA specific exonuclease that digests DNA in the 5' → 3' direction. Unprotected single-stranded DNA from step 4 is largely destroyed by this step.

Resin-bound DNAs were separated from residual reaction buffers by centrifugation through Corning Costar spin-X centrifuge tube filters (Sigma-Aldrich), followed by two washes in ice cold TBS. DNA-protein complexes were eluted from resin using FLAG peptide (Sigma-Aldrich) at a final concentration of 150 ng/μl in 100 μl TBS, with slow shaking for 30 minutes. Supernatants were collected by centrifugation through the spin-X columns and saved as immunoprecipitated (IP) samples. To reverse formaldehyde crosslinking, IP samples were pre-incubated with 90 μl ChIP elution buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS) and 10 μl of protease (Sigma-Aldrich) (40 mg/ml in TBS) for 2 hours at 42^{0}C, followed by 6 hours at 65^{0}C. Qiagen PCR-purification spin columns were used to purify DNAs.

To prepare ChIP-exo libraries for high-throughput sequencing, samples were incubated at 95^0C for 5 minutes to denature double stranded DNAs. Primer P2 (5 pmol, see Table 2.2) was added, allowed to anneal for two minutes at 30^0C, and extended using 10 U of phi29 polymerase at 30^0C for 20 minutes. The polymerase was inactivated by incubation at 65^0C for 10 minutes. Blunt-end ligation of the P1 adaptor (10 uM) was accomplished using T4 DNA ligase 1000U at 25^0C for 60 min, followed by incubation at 65^0C for 10 minutes to inactivate the ligase. DNA was purified using Agencourt AMPure magnetic beads and amplified by PCR using DNA polymerase Phusion in 18 PCR cycles. Finally, samples were purified using 80 µl Agencourt AMPure magnetic beads (Beckman Coulter Genomics), following the manufacturer's instructions, and eluted to a final volume of 30 µl.

RNA isolation and preparation for RNA-Seq

Total RNA was prepared using an RNeasy Kit (Qiagen) following the manufacturer's instructions, using the optional on-column DNaseI digestion. RNA was treated twice with DNase I (Ambion) to remove residual DNA and then cleaned and concentrated using RNA cleanup and concentrator-5 (Zymo Research). Integrity of the RNA was assessed using the Agilent Bioanalyzer (Cornell University Life Sciences Core Laboratory Center Microarray Facility, Cornell University).

Depletion of processed RNAs and ligation of tag. Ribo-Zero™ rRNA Removal Kit (Epicenter) was first used to remove ribosomal RNA (rRNA). RNAs were then treated with Terminator™ 5´-Phosphate-Dependent Exonuclease (Epicentre), as described previously [62], to digest RNAs terminated by a 5´-monophosphate group, leaving RNAs terminated by 5´-triphosphate or 5´-hydroxyl groups undigested. Tobacco acid pyrophosphatase (TAP,

Epicentre Biotechnologies) was then used to convert terminal 5'-triphosphate moieties to 5'-monophosphate. The 3'-ends were blocked by treatment with $NaIO_4$ to prevent circularization before ligation of an RNA oligonucleotide (5'- ACA UCC ACA UCC UAG UAC -3'; IDT custom RNA oligo) to RNA 5'-ends. The ligation reaction was incubated overnight at 16^0C. Products were recovered using RNA cleanup and concentrator-5 (Zymo research) and eluted in 10 ml H_2O.

Construction of cDNA libraries for 5' mapping and RNAseq. Stranded RNA-Seq libraries were prepared using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre) and 500 pg to 50 ng rRNA-depleted RNAs following the manufacturer's protocol. Briefly, RNAs were fragmented and cDNA synthesis was carried out using random-sequence primers containing a tagging sequence at their 5'-ends. The 3'-ends were tagged using the Terminal-Tagging Oligo. Di-tagged cDNAs were purified using magnetic beads and subjected to a limited-cycle PCR as recommended (10-15 cycles). Samples were indexed for multiplexing using the appropriate primers. Libraries were purified using the AMPure XP system (Beckman Coulter) in 20 μl total volume. Library quality was assessed using the Agilent Bioanalyzer. Sequencing was performed on the Illumina HiSEQ2000 by the Cornell DNA Sequencing Core Facility.

Alignment of sequence reads to the DC3000 genome sequence and profile generation

For RNA-Seq data, reads bearing the unique 18 nucleotide sequence (tag) at the 5' end were first identified and separated from the set of all RNA-Seq reads. The tag was then removed from each read, leaving the nucleotide sequences derived from the original RNAs. Thereafter, the de-tagged RNA-Seq reads (5' capture data), the untagged RNA-Seq reads

and reads obtained from the ChIP-exo procedure (ChIP-Seq data) were handled in the same way. Quality scores for the sequence reads were accessed using FastQC [150]. The first 75 nucleotides of each read with quality score of 20 or above (99% of inferred base calls are accurate) were aligned to the reference chromosome of DC3000 (accession number: AE016853), plasmid A (accession number: AE016855) and plasmid B (accession number: AE016854) using SOAPalign/soap2 [151].

Reads that aligned perfectly to a single location were retained and all others were discarded. The "sinister profiles" are histograms representing the number of trimmed reads whose 5'-ends uniquely map to each position [131]. A profile has values for each of the two strands of the genome. Profiles were visualized using the Artemis genome viewer [152] as previously described [131]. Profiles for the main chromosome and plasmids are available in Supplemental data.

Identification of regions enriched by ChIP

Enriched regions (those overrepresented in the ChIP-Seq data set) were identified using Genetrack [153]. The sinister profiles were formatted to a GeneTrack compatible format using a custom Python script. The GeneTrack analysis merged signals in cases where 5 adjacent positions had aligned read counts of 1000 or more at each position. Signals were constrained to be at least 700 bps apart from each other. Finally, overlapping signals on opposite strands were combined to generate ChIP-Seq 'peaks' (candidate HrpL-binding sites). Any signal mapping to one strand only (i.e., without an accompanying signal on the other strand) was discarded.

The normalized number of sequence reads associated with each site enriched by ChIP-Seq was also computed. The Un-normalized 'peak' height is an average of reads on both chromosomal strands in a window of 30 bps centered at ChIP-Seq 'peak'. The background height was average of reads on both chromosomal strands in 2 windows of 30bps that are 400 bps upstream and downstream of the ChIP-Seq 'peak'. This background height was used to normalize the corresponding un-normalized 'peak' height. The normalized value on *ΔhrpL* mutant was also computed to present enrichment attributable to unspecific binding.

Motif detection using MEME

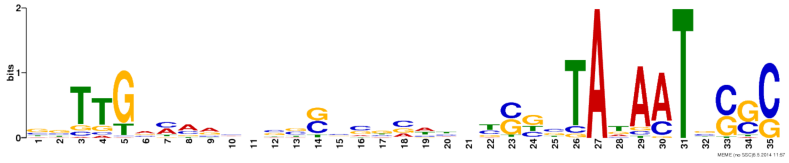
Sequences of length 40-50 bps (FASTA format) upstream from captured 5' transcription start sites were used as input to MEME [154] with the following parameters:

-dna -mod anr –nmotifs 20 -minw 14 -maxw 35 -maxsize 150000

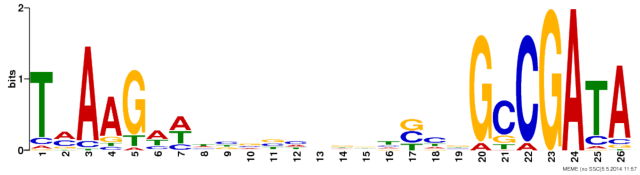
The MEME package also generates sequence logos for each sequence pattern detected (Figure 2.1).

From *hrpL-FLAG*

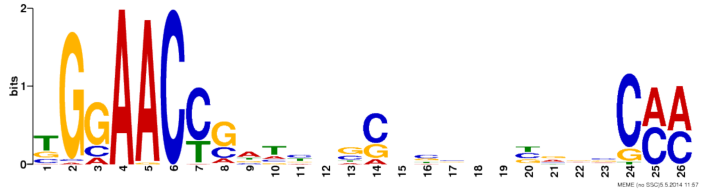
Motif 1



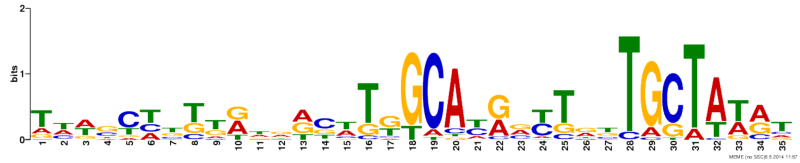
Motif 2



Motif 3

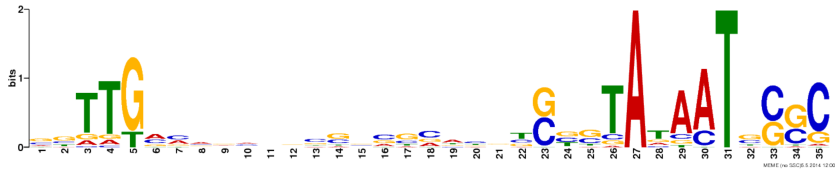


Motif 4



From *ΔhrpL*

Motif 1



Motif 2

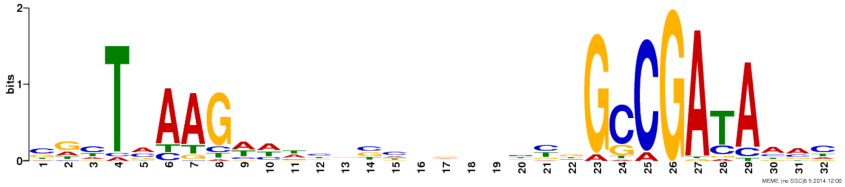


Figure 2. 1 Likely promoter motifs recovered by MEME using 5'-end capture data from *hrpL-FLAG* and *ΔhrpL* cells

HrpL regulon orthologs and matches to *hrp* promoter motifs

We first constructed a profile hidden Markov model (HMM) [155] using confirmed *hrp* promoters as a training set (Table 2.3). DNA sequences for 1060 closed and draft *Pseudomonadales* genomes were obtained from NCBI (Table 2.4†). Prodigal [156] was used to generate uniform gene predictions for all genomes. Next, BLASTP [157] was used with an e-value threshold of 1e-6 to compare DC3000 protein sequences to proteins encoded in other genomes, and the reciprocal best matches were retained as presumptive orthologs. DNA sequences upstream from each DC3000 HrpL regulon ortholog were then extracted, with sequence length adjusted to account for the distance between the DC3000 promoter and its closest downstream gene (100 bps to 1500 bps), except for PSPTO_4750 (antisense sequence was extracted) and PSPTO_4955 (sequence in the middle of the gene was extracted). The sampled sequences were scanned with the profile HMM. In each case, the best scoring motif match in each upstream sample was noted as a potential *hrp* promoter (Table 2.4). The same HMM was used to scan the DC3000 chromosome to determine if additional potential *hrp* promoters could be associated with weak ChIP-Seq signals (Table 2.5). Custom scripts were used as necessary to simplify intermediate steps in the analysis.

† Table 2.4 is substantially large with more than 1000 lines. It is not shown here. For reference, refer to the published paper associated with this work.

Table 2. 3 *hrp* promoter motif sequences included in the HMM training set.

>61504:61536 REVERSE
GGAACTGCAAGCTGGTGCGGCTCCACTAA
>82447:82478 FORWARD
GGAACTCATCACCGCGAATCG-CCACTCA
>404752:404784 FORWARD
GGAACTGCAACGTTGTTTCGG-TCACTCA
>522444:522475 REVERSE
GGAACCGAATCCATATTTCGA-CCACCCA
>550602:550634 REVERSE
GGAACCTGATGCTGCTCAGTGACCACTCA
>572473:572504 FORWARD
GGAACCCATAGAGCCTGCCTG-CCACTTA
>648424:648456 REVERSE
GGAACCGAATCCATCTCGAGGGCCACTCA
>649735:649766 REVERSE
GGAACTGAACCGCTTATGAAA-CCACTCA
>905339:905371 FORWARD
GGAACCAAAACTGGAAAAACATCCACTCA
>921879:921911 REVERSE
GGAACCTCACGCTTAGTGATGACCACGCA
>939413:939445 REVERSE
GGACCCGAATCCGTCTTAAACACCACTCA
>941100:941132 FORWARD
GGAACCGATAACGACTTTTTGGCCACTTA
>946154:946185 FORWARD
GGAACCCAAGAGCCCTTGCGA-CCACACA
>949826:949858 REVERSE
GAAACCGAAACGGCGTTGCTTGCCACACA
>954203:954234 FORWARD
GGAACCGATCCGGTTGCCTGG-CCACTCA
>981177:981209 FORWARD
GGAACCCGATGACACAAGGCG-ACACTCA
>1116378:1116410 REVERSE
GGAACCACATCATGGGTAAAAGCCACGAA
>1504886:1504917 FORWARD
GGAACCGCATCACGTCTTGAA-CCACAGA
>1505219:1505251 FORWARD
GGAACTGTTGTTGCGCAGTGT-GCACTCA
>1507652:1507684 FORWARD
GGAACCGTCAACCGATCCGGGACCACACA
>1510785:1510816 REVERSE
GGAACCGGTCGCTGCGCTTTG-CCACTCA
>1510881:1510913 FORWARD
GGAACTGAAATGCCTATGCCTGCGACTCA
>1519570:1519601 REVERSE

GGAACCCGCTGGCATTGCATG-CCACTCA
>1519666:1519697 FORWARD
GGAACCGTAACGGCGAGCGTG-CCACGTA
>1524204:1524236 FORWARD
GGAACCGATTCGCAGGCTGCTGCCACCTA
>1528184:1528216 FORWARD
GGAACCGCTCGGCGGGTTTGCTCCACTCA
>1536874:1536905 REVERSE
GGAACTGAAATCGATGCTCGA-CCACTTA
>1542621:1542653 REVERSE
GGAACTGATCCGGGACCGTGACCCACTCA
>1543416:1543447 FORWARD
GGAACCAACTTGCACCTTCAA-CCACACA
>1548389:1548420 FORWARD
GGAACCGATTTCGATGAGTCG-CCACACA
>1731421:1731453 REVERSE
GGAACCACTGAAGAGTTTTAAGCCACTCA
>2279883:2279915 FORWARD
GGAACCCTGCGCAGGTCATTGACCACTCA
>2973825:2973856 FORWARD
GGAACCGAGTCACTCA-GTGAACCACTCA
>3470185:3470217 REVERSE
GGAACTCTTTCCTGCTCTTTTGCCACACA
>4515296:4515328 REVERSE
GGAACCGATCCGCTCCCTATGACCACTCA
>4621129:4621160 FORWARD
GGAACTCTTTCCCTGCGCTTT-CCACTCA
>4881097:4881129 FORWARD
GGAACCGAATCCGCCTCAAAGTCCACACA
>5186123:5186154 REVERSE
GGAACCCCTGCGCGTCCAGCG-CCACTCA
>5192613:5192644 REVERSE
GGAACCGGACGAGGCTTTTTA-CCACTCA
>5305220:5305252 FORWARD
GGAACCTTTTTCTTGCATAGCGCCACAGA
>5330688:5330720 FORWARD
GGAACTGCAAGCTGGTGCGGCTCCACTAA
>5344375:5344407 REVERSE
GGAACCGTTAACGGCCCAGCGACTACACA
>5348578:5348610 REVERSE
GGAACCCCACAAGCACTTATGACCACGCA
>5350034:5350065 FORWARD
GGAACCCACGAGCCCTTGTGA-CCACATA
>5355224:5355256 REVERSE
GGAACTCTACGCTTGGCGATGACCACGTA
>5355530:5355562 FORWARD
GGAACTATCTTCCCACACGGAGCCACTTA

>5361707:5361739 REVERSE
GGAACCAAATCCGCCTCAAAGTCCACTCA
>5418197:5418228 REVERSE
GGAACCAGATCTCGTTGCTTG-CCACCAA
>6085756:6085788 FORWARD
GGAACCGCCTCGAGCAGAGGCTCCACTCA
>15821:15849 FORWARD
GGAACTGCCGGCTGGTGCGGCTCCACTAA
>406210:406238 FORWARD
GGAACCGATCTCGCCATCGAGGTAACTCA
>880830:880858 FORWARD
GGAACCACCTATTTCACGTCGACAACCAA
>922925:922953 REVERSE
GGAACCAAATTGCTGGCTACGCTAACCAA
>939675:939703 FORWARD
GGAACCCATCAGCACAAGCGCACCACGAA
>1115513:1115541 REVERSE
GGAACTAAATAATTTCCTAGG-CGACTAA
>1802305:1802333 FORWARD
GGAACTGCGCCGCGACCTCAAACCACAAA
>2012108:2012136 FORWARD
GCAACCAAACTGAACCATTCGGCCACTCA
>2304331:2304359 REVERSE
GGAACTGAGTGACGGAAACGTTTCACTAA
>2984435:2984463 FORWARD
GGAACTTTCGGTACAGGAGCTACCACAAA
>2990249:2990277 REVERSE
GGAACTGTTTTCTGGCAATGGCCCACTGA
>3768950:3768978 REVERSE
GGAACTGTTGGTCAAGCGTTTGCCACTGA
>3929005:3929032 FORWARD
GGAACT-GATTCCCAGGATCGGTGACTAA
>4199604:4199632 FORWARD
GGAACTGGTCGGCATCCCTGAGCCACATA
>4457004:4457031 REVERSE
GGAAG-GACATCCTCTTCGTGATCACCTG
>4895201:4895228 REVERSE
GGAACT-TCAAGGCTGACGCGGTAACTCA
>5328022:5328050 REVERSE
GGAACCGATAACAAATTTTTGGCCACATA
>5346761:5346788 REVERSE
GGAAC-GTGTCCGCTGTTACGGCCACTTA
>5384480:5384507 FORWARD
GGAACC-CATGGAGCTTGTTAGCCACATA
>5616671:5616699 FORWARD
GGAACCATTGAGCCTGTTCGCGCCACTGA
>5751475:5751504 FORWARD

GGAAACCGGACAGCAGTTGCTGGCCACTC
>5960164:5960192 REVERSE
GGATCAAAAATGGGCCTGGATGCCACTAA

Table 2. 4 Orthologs and *hrp* promoter motif scores for HrpL regulon members in 1060 *Pseudomonadales* genomes (see note).

Table 2. 5 Additional *hrp* promoter-like sequences with limited experimental support.

Operon: PSPTO identifier for operon/gene immediately downstream from *hrp* promoter like sequence.
*1654: predicted *hrp* promoter is antisense relative to this gene.
Function: annotated function for identifying gene.
Sequence: potential –35 and -10 regions are underlined and aligned
hrp promoter E-value: E-value for match to profile HMM.
Coordinate: DC3000 genome coordinate for *hrp* promoter-like sequence
ChIP-Seq: qualitative assessment of HrpL-binding activity.
RNA-Seq: "low" for low abundance transcripts (less than 100 reads in the coding region), "x" for no reads.
GFP/OD: reporter fluorescence normalized to OD_{600} for reporter constructs cultured in KB or *hrp*-minimal media. ND indicates not determined.

Operon	Function	*hrp* promoter like sequence	*hrp* promoter E-value	coordinate	ChIP-Seq	RNA-Seq	gfp/od on KB	gfp/od on hmm
0816	type IV pilus biogenesis protein	GGAACCACCTATTTC ACGTCGACAACCAA	0.0031	880830..880858	weak	low	16.6 ± 18.4	180.3 ± 28.7
1654*	conserved protein of unknown function	GAAACCGCAAGATCA AACGGCACCACCCA	0.0055	1819661..1819689	weak	x	ND	ND
3019	oxidoreductase, Gfo/Idh/MocA family	GGAAC- GCGACTATGCTGCGG GCCACTCA	0.0024	3393411..3393438	weak	low	ND	ND
3122	6-phosphogluconate dehydrogenase	GGAACCGACGTTCTG CCTCGCTTGACTAA	0.0072	3508787..3508815	weak	low	ND	ND
4580	conserved protein of unknown function	GGAACTGGTGCAGGG CGCTT-GCCACGAA	0.0013	5176134..5176161	weak	low	ND	ND

qRT-PCR and qPCR

cDNA synthesis was accomplished using qScript cDNA Synthesis (Quanta, Biosciences) and random primers as part of the kit. qPCR steps were performed using iQ SYBR Green Supermix (Bio-Rad). Primer pairs (see supplemental data) were selected to amplify a region of approximately 100 bps (Beacon Designer™). For evaluating enrichment of ChIP-Seq binding sites, the primers amplified within the putative enriched regions. Enrichment (fold change) was calculated for DNA recovered at each tested binding site in the immune-precipitated (IP) sample compared to that in the lysate sample. The housekeeping genes, *gyrA* and *gap1*, were the internal and negative controls, respectively, for all tested regions. To test for transcript abundance, regions downstream of mapped transcriptional start sites were amplified. Transcript levels for each region (including the negative control, *gap1*) were calculated relative to the level for the housekeeping gene *gyrA*.

Plant virulence assays

Solanum lycopersicum or *Nicotiana benthamiana* plants were germinated and grown in a greenhouse with approximate 16/8 hr. light/dark cycles. Four to five week old tomato plants or two to three week old *N. benthamiana* plants were inoculated with a 3 x 10^4 CFU/ml bacterial suspension using blunt syringe infiltration. Bacteria were recovered from plants by sampling leaf tissue at the site of infection using a #2 disk punch (3 disks, total area 0.589 cm^2) at 2 days post infection (dpi), 4 dpi and 6 dpi. Leaf disks were homogenized by mechanical disruption in 700 ml of 10 mM $MgCl_2$. Serial dilutions of the

tissue homogenate were plated on KB agar supplemented with rifampicin and the number of colony forming units per milligram leaf tissue was calculated.

Hypersensitive Response (HR) assays

Inoculation was performed as described above for virulence assays except the inoculum concentration was 3 x 10^7 CFU/ml for DC3000 derivatives unless otherwise noted. HR was observed and documented by photography after 2 days.

Cya translocation reporter assays

Translocation assays were performed as described previously [139]. DC3000 strains were grown overnight at 28^0C on KB agar medium with appropriate antibiotics, resuspended in 10 mM $MgCl_2$, and adjusted to an OD_{600} of 0.05 (~5x10^7 CFU/ml). Bacterial suspensions were inoculated into *N. benthamiana* leaves using a blunt tip 1 ml syringe and plants were placed on the lab bench. After 6 hours, 2 leaf discs per sample were excised using a 1 cm-diameter cork borer and frozen in liquid nitrogen along with 2 copper-coated beads (Copperhead BBS, Crosman Corporation). The leaf discs were finely ground by vigorous vortexing and resuspended in 300 µl of 0.1 M HCl. The leaf extract was centrifuged at 6500 rpm for 10 minutes, and the supernatant was transferred to a fresh tube. A 10-fold dilution was prepared in 0.1 M HCl and used for the Cya assay since it was observed that diluting the extract resulted in greater sensitivity. Total pmol cAMP in each sample was determined using the Direct cAMP ELISA kit (Enzo Life Sciences) following the manufacturer's instructions. Three plants were inoculated in each experiment and the samples were analyzed in 2-5 independent experiments.

Results

Candidate *hrp* promoter regions identified by ChIP-Seq

As a direct regulator of the HrpL regulon, the sigma factor HrpL (PSPTO_1404) is required for DC3000 virulence. To perform a global inventory of genomic sites likely to bind this sigma factor, we tagged HrpL with a FLAG epitope at the C-terminus by modifying the *hrpL* locus at its native position in DC3000 chromosome (Figure 2.2A) and performed a ChIP-Seq analysis. We first confirmed that cells bearing the tagged protein retained the ability to stimulate the hypersensitive response in a plant assay (Figure 2.2B) and established that HrpL-FLAG retained its ability to support transcription from a known HrpL-responsive promoter (Figure 2.2C). Results from a pilot experiment suggested that samples for ChIP-Seq should be harvested at 1.5 hours after a shift from non-inducing to *hrp*-inducing conditions, given the relatively high abundance of transcripts for *hrpL* and *hopQ1-1*, a confirmed effector, at this timepoint (Figures 2.2D-E).

Using the high-resolution ChIP-exo procedure described by Rhee *et al.* [148] we prepared samples for high-throughput sequencing. Table 2.6 summarizes the reads and mapping statistics of the two libraries using data generated by the Illumina Genome Analyzer. As a control, ChIP-Seq samples for the Δ*hrpL* strain were prepared using a conventional ChIP-Seq protocol [149] that omits the steps in ChIP-exo that destroy most of unbound DNA [148] (described in Materials and Methods). The read counts from this sample were compared to those from the *HrpL-FLAG* sample to determine levels attributable to non-specific binding.

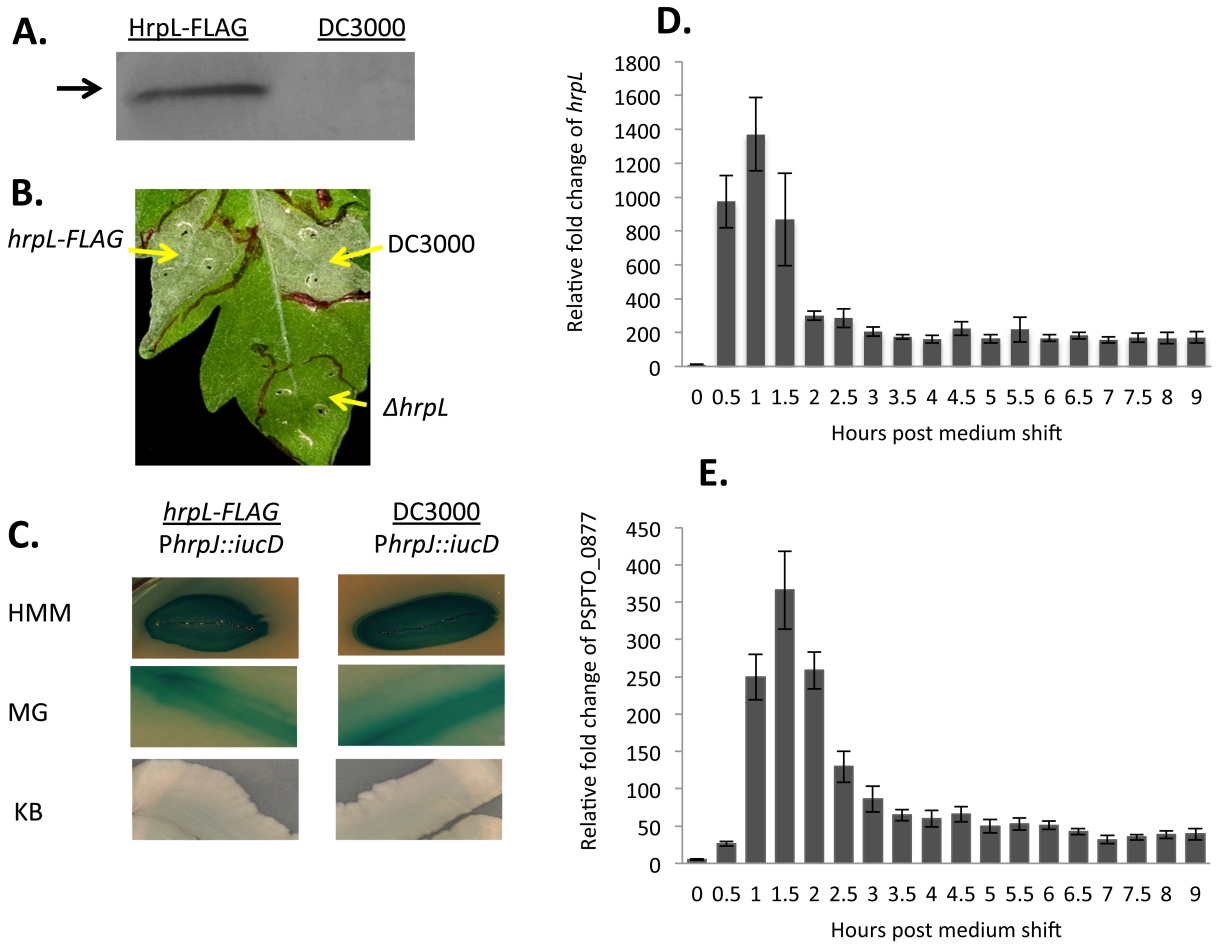


Figure 2. 2 HrpL-FLAG is functional.

(A) HrpL-FLAG is recognized by anti-FLAG antibody in a Western Blot. (B) *hrpL-FLAG* and WT DC3000 strains evoke the hypersensitive response while it is abolished in the *ΔhrpL* strain. Bacteria were infiltrated via a blunt syringe into three independent leaves at $3x10^8$ CFU/ml. Photos were taken after 3 days. Symptoms are identical among three replicates. (C) Expression of a plasmid-based (pBS181) β-glucuronidase (GUS) reporter driven by a *hrp* promoter (*hrpJ*) in *hrpL-FLAG* and WT DC3000 backgrounds. Three different media were used: HMM (*hrp*-minimal medium for highest induction of *hrp* promoters), MG (Mannitol-Glutamate medium for intermediate induction of *hrp* promoters but better bacterial growth), and KB (King's B rich medium for repression of HrpL regulon expression). All plates contain appropriate antibiotics to maintain plasmids and X-Gluc, a substrate of GUS. (D-E) Relative fold change of transcript levels for *hrpL* and *hopQ1-1* (PSPTO_0877) after medium shift from KB to MG (supplemented with ferric iron at 50 μM final concentration) over 9 hours.

Table 2. 6 Read mapping statistics

Sample	Sequence read type	*hrpL-FLAG*	*ΔhrpL*
ChIP-Seq	Total reads	45,479,930	22,170,715
	Reads that align uniquely to the chromosome	35,185,956	9,513,703
	Reads that align uniquely to plasmid B	2,699,119	571,096
	Reads that align uniquely to plasmid A	3,308,504	573,047
RNA-Seq	Total reads	36,272,771	44,500,185
	Reads containing 18-mer	356,915	3,351,960
	Untagged reads that align uniquely to the chromosome	29,349,347	32,953,858
	Untagged reads that align uniquely to plasmid B	1,426,734	515,184
	Untagged reads that align uniquely to plasmid A	878,164	426,741
	Tagged reads that align uniquely to the chromosome	237,794	2,069,913
	Tagged reads that align uniquely to plasmid B	20,570	30,903
	Tagged reads that align uniquely to plasmid A	7,452	30,501

Using the Genetrack [153] analysis package, we identified 73 enriched regions in the DC3000 genome, of which 52 can be associated with one of the 54 annotated *hrp* promoters (Table 2.7), and 21 appear in regions without known *hrp* promoters (Table 2.8). The two annotated promoters that were negative with respect to enrichment were PSPTO_1370 (type III effector HopN1) and PSPTO_3489 (a sugar ABC transporter/ATP-binding protein). Visual inspection of these regions in the ChIP-Seq profile also showed no evidence of enrichment. Because PSPTO_5633 and PSPTO_B0003 have identical coding and upstream sequences, we counted them as a single enrichment instance, reducing the number of novel regions identified by ChIP-Seq to 20. Among these promoter candidates, only two are obviously associated with genes that encode proteins involved in the T3SS, namely PSPTO_4721 (type III chaperone ShcV) and PSPTO_5618 (pseudogene for type III effector HopAT1). The remaining candidates are upstream of genes encoding enzymes, hypothetical proteins or proteins with other functions.

Table 2. 7 Data for all *hrp* promoters

Operon: PSPTO identifier for genes immediately downstream from promoter. Identifiers in parentheses indicate that the operon contains multiple genes. Operons are adapted from Table 2 in Ferreira *et al.* [65]
Function: annotated function for operon-identifying genes.
Coordinate: DC3000 genome coordinate for the region bracketing the –35 and –10 regions of the promoter. "c" designates that the promoter is found on the complementary strand.
Evidence: Experimental evidence for *hrp* promoters from this study.
H: *hrp* promoter motif found;
B: Binding activity for HrpL (ChIP-qPCR and/or ChIP-Seq) observed;
(*) ChIP-Seq enrichment occurs between two divergent promoters. The asterisk designates the promoter located further from the enriched area.
I: Induction observed in promoter fusion (threshold = 2.4; 2 x negative control);
Q: HrpL dependent transcription of regions downstream from *hrp* promoters in DC3000 in qRT-PCR.
S: mRNA 5'-end captured (TSS). Absolute values for read counts appear in [b] and [d].
[a] Distance between *hrp* promoters (3'-end of –10 region) and captured 5'-end within 100 bps using data from this study. The first value is the distance to the closest captured end. The second value is the distance to the captured end with the most sequence read counts within 100 bps from the *hrp* promoter.
[b] Read counts from 5'-end capture data within 100bps using data from this study. The first value is read count for the closest captured 5'-end to the *hrp* promoter. The second value is read count for the captured end with the most sequence reads within 100 bps from the *hrp* promoter. Values are based on one replicate as described in the text and methods.
Published data:
[c] Distance between *hrp* promoter (5'-end of –10 region) and captured 5'-end within 100 bps using data from Filiatrault *et al.* [62]. The first value is the distance to the closest captured 5'-end. The second value is the distance to the captured end with the most sequence read counts within 100 bps from the *hrp* promoter.
[d] Read counts for captured 5'-ends within 100 bps using data from Filiatrault *et al.* [62]. The first value is the read count for the closest captured 5'-end to the *hrp* promoter. The second value is the read count for the captured end with the most sequence read counts within 100 bps from the *hrp* promoter.
• Promoters and genes: promoters and differentially expressed genes associated with the HrpL regulon reported elsewhere.
- Y: Mucyn *et al*. [130]
- C: Chang *et al*. [85]
- F: Fouts *et al*. [80]

- M: Ferreira *et al.* [65]exhibited differential expression in microarray (M); exhibited differential expression in microarray and/or RT-PCR but did not pass other criteria (M*); genes present on microarray but no evidence for differential expression (M_o); genes not on microarray and not tested by RT-PCR (M_1); evidence for existence of peptide fragment detected by mass spectrometry (M_2). The score for the match between the hidden Markov model and the *hrp* promoter reported by Ferreira *et al.*

Background colors:

BLUE: experimental evidence for HrpL-dependent expression as reported in [65], and [80, 85]. The first study used microarray to compare a *hrpL* deletion with wild-type DC3000 in *hrp* minimal medium. The second and third studies compare a *hrpL* deletion to a strain in which HrpL was overexpressed.

GREEN: experimental evidence reported in [65], but not in [80, 85].

ORANGE: experimental evidence reported in [80, 85], but not in [65].

WHITE: experimental evidence for HrpL-dependent expression only in a recent RNA-Seq by Mucyn *et al*. [130].

GRAY: no previously experimental evidence for HrpL-dependent expression.

Operon	Function	*hrp* promoter coordinate	Evidence	Distance to TSS[a]	TSS[b] Counts	Published data		
						Distance to TSS[c]	TSS[d] Counts	Promoters and genes
44	Type III effector HopK1	61504..61536	HBQS	51, 81	4, 4634	1, 8	1, 47	CM* 13.2
61	Type III effector HopY1	82447..82478	HBQS	4, 6	101, 597	4, 5	99, 965	FCM 16.3 Y
370	MATE efflux family protein	404752..404784	HBS	3, 6	4, 2725	6, 6	390, 390	FCM_0 13.2 Y
371	*iaaL* indoleacetate-lysine ligase	406210..406238	HBIQS	2, 6	4, 9483	2, 6	1, 2452	FCM_0 10 Y
503-(502-501)	Type III chaperone protein ShcF-(hopF2-hopU1)	550602..550634	HBQS	77, 79	4, 42	6, 6	67, 67	M-FCM_0-M_0 18.8 Y
524	M20/M25/M40 family peptidase	572473..572504	HBS	4, 7	4, 509	5, 7	22, 34	M* 18.3 Y
588	Type III effector HopH1	648424..648456	HBQS	22, 80	4, 736	6, 7	22, 598	CM 26.1 Y
589	Type III effector HopC1	649735..649766	HBQS	60, 80	4, 1220	7, 7	120, 120	FCM 20.4 Y
834-(835-836-837-838)	alcohol dehydrogenase, zinc-containing-(ribD C-terminal domain protein-conserved domain protein-conserved protein of unknown function-major facilitator family transporter)	905339..905371	HBQS	2, 6	4, 3621	4, 6	11, 115	CM*-CM_0-CM_0-M_0- M_0 19.2 Y
852	Type III effector HopAJ1	921879..921911	HBQS	42, 81	8, 1518	6, 8	4, 762	FCM 15.5 Y
871	macrolide efflux protein, putative	939675..939703	HBIQS	4, 6	122, 643	4, 6	17, 51	Y
873	amidinotransferase family protein	941100..941132	HBQS	2, 4	147, 774	2, 6	32, 294	M_0 19.2 Y
876	Type III effector HopD1	946154..946185	HBQS	4, 5	4, 206	2, 5	1, 75	FCM 20.4 Y

877	Type III effector HopQ1-1	949826..949858	HBQS	5, 80	17, 6590	5, 7	1, 559	CM 23.3 Y
883	Type III effector HopR1	954203..954234	HBQS	4, 7	29, 833	2, 7	1, 2361	FCM 20.7 Y
1022	Type III effector HopAM1-1	1116378..1116410	HBQS	43, 79	4, 6838	x	x	FCM 18.4
1369	Type III chaperone protein ShcN	1504886..1504917	HBQS	2, 5	164, 2826	2, 5	10, 205	FM 18.4 Y
1370	Type III effector HopN1	1505219..1505251	QS	5, 6	42, 307	0, 6	5, 34	FCM 12.8 Y
1372	Type III effector HopAA1-1	1507652..1507684	HBQS	3, 3	42, 42	3, 3	7, 7	FCM 18.7 Y
1373	Type III helper protein HrpW1	1510785..1510816	HBQS	77, 80	4, 307	3, 7	3, 670	FCM 15.4 Y
1374-(1375-1376)	type III chaperone ShcM-(hopM1-shcE)	1510881..1510913	HBQS	3, 6	4, 458	4, 6	54, 114	CM 17.9 Y
1377	Type III effector protein AvrE1	1519570..1519601	HBQS	67, 80	4, 967	6, 7	1, 238	CM_0 15.2 Y
1378	membrane-bound lytic murein transglycosylase D	1519666..1519697	HBQS	4, 6	50, 122	4, 6	45, 67	FCM 20.7 Y
1381-(1382-1383-1384-1385-1386)	Type III helper protein HrpA1-(hrpz1-hrpB-hrcJ-hrpD-hrpE)	1524204..1524236	HBQS	3, 6	29, 58959	3, 6	2, 3307	CM-CM-M-M-M-M 20.6 Y
1387-(1388-1389-1390-1391)	Type III secretion protein HrpF-(hrpG-hrcC-hrpT-hrpV)	1528184..1528216	HBQS	5, 6	21, 1350	5, 6	6, 654	CM_1-M-M- M_1-M 20.4 Y
1398-(1397-1396-1395-1394-1393-1392)	Type III secretion protein HrpP-(hrcQa-hrcQb-hrcR-hrcS-hrcT-hrcU)	1536874..1536905	HBQS	72, 80	4, 341	5, 6	12, 255	M-M-M-M- M_1-M_0-M_0 20.4 Y
1403-(1402-1401-1400-1399)	Type III secretion protein HrpJ-(hrcV-hrpQ-hrcN-hrpO)	1542621..1542653	HBQS	19, 80	4, 282	5, 7	2, 60	CM-M-M-M_0-M 20.4 Y
1405-(1406)	Type III helper protein HrpK1-(HopB1)	1543416..1543447	HBQS	6, 6	151, 151	5, 6	2, 388	CM 16.4 Y

1645	MarR family transcriptional regulator	1802305..1802333	HBIQS	5, 7	269, 408	3, 5	1, 41	M* 12.3
1843	aspartate kinase	2012108..2012136	HBIQS	5, 6	168, 1707	0, 6	1, 116	Y
2105	thiamine biosynthesis lipoprotein, putative	2279883..2279915	HBS	1, 6	147, 7023	1, 6	1, 91	FCM* 22.8 Y
2130	LuxR family DNA-binding response regulator	c(2304331..2304359)	HBIQS	79, 80	21, 185	7, 7	32, 32	M_0 11.3 Y
2678-(2679)	Type III helper protein HopP1-(protein of unknown function)	2973825..2973856	HBQS	0, 3	21, 400	1, 3	4, 25	FCM 19.6 Y
2691	TerC family membrane protein	2984435..2984463	HBIQS	5, 5	67, 67	4, 5	1, 23	M* Y
2696	mutT/nudix family protein	c(2990249..2990277)	HBIQS	60, 79	4, 980	6, 17	31, 160	M_0 11.6
3087	Type III effector HopAB2	3470185..3470217	HBS	37, 79	4, 4735	6, 6	908, 908	FCM 21.2 Y
3331	protease inhibitor Inh	c(3768950..3768978)	HBIQS	78, 80	8, 59	0, 22	1, 709	
3481	Hypothetical protein	3929005..3929032	HBQS	4, 5	17, 841	2, 5	1, 810	
3489-(3488)	sugar ABC transporter, ATP-binding protein-(sugar ABC transporter, permease protein)			x	x	x	x	M-M 11.6
3721	fabI enoyl-(acyl-carrier-protein) reductase	4199604..4199632	HBIS	6, 6	84, 84	1, 6	1, 3	M_0 11
3948_49	Hypothetical protein	c(4457004..4457031)	HBIQS	30, 30	8, 8	28, 70	1, 453	
4001	Type III effector protein AvrPto1	4515296..4515328	HBQS	27, 80	4, 711	4, 7	4, 1194	FCM 22.3 Y
4101	Type III effector HopAK1	4621129..4621160	HBQS	1, 7	4, 147	5, 7	22, 253	FCM 15.1

4331	Type III effector HopE1	4881097..4881129	HBQS	4, 7	4, 4088	5, 7	12, 348	CM 23.6 Y
4340	insecticidal toxin protein, putative	c(4895201..4895228)	HBIQS	80, 80	25, 25	5, 7	1, 3	
4589-(4588)	Type III chaperone ShcS2-(hopS2)	5186123..5186154	HBQS	30, 79	4, 2031	5, 6	1, 223	CM_0-CM_1 18.7 Y
4599-(4597)	Type III chaperone ShcS1-(hopS1)	5192613..5192644	HBQS	43, 80	4, 892	2, 7	1, 1277	FM-CM 24.2 Y
4691	Type III effector HopAD1	5305220..5305252	HBS	3, 5	4, 971	4, 5	3, 55	CM_0 8.7 Y
4699	non-ribosomal peptide synthetase, terminal	c(5328022..5328050)	HBIQS	19, 77	4, 235	2, 4	9, 25	M_0 16.8
4703-(4704-4705)	Type III effector HopAQ1-(corR-corS)	5330688..5330720	HBS	3, 7	50, 122	1, 7	1, 16	CM_1 13.2 Y
4718	Type III effector HopAA1-2	5344375..5344407	HBQS	36, 81	4, 3427	1, 8	1, 493	M^* 18.8 Y
4721	Type III chaperone ShcV	c(5346761..5346788)	HBIQS	41, 79	4, 261	3, 74	1, 25	M_0 12.2 Y
4722	Type III effector HopAO1	5348578..5348610	HBQS	1, 79	8, 38	5, 6	2, 17	FCM_0 13.8 Y
4724-(4725)	Type III effector HopD-(IS52, transposase)	5350034..5350065	HBQS	4, 5	4, 67	4, 61	8, 410	M_1 16
4727	Type III effector HopG1	5355224..5355256	HBQS	74, 81	4, 63	6, 8	1, 10	CM 14.4 Y
4733	Hypothetical protein	5361707..5361739	HBS	60, 79	4, 400	6, 8	174, 239	M* 22.2 Y
4750	Hypothetical protein	5384480..5384507	HBIQS	4, 6	59, 261	4, 6	3, 28	
4776	Type III effector HopI1	541819.. 5418228	HBQS	76, 80	4, 908	5, 7	13, 128	FCR 16.5 Y
4955	bifunctional thiosulfate	5616671..5616699	HBIQS	3, 7	412, 421	0, 3	6, 254	M_0 10.3
5053	Hypothetical protein	5751475..5751504	HBIQS	4, 5	4, 1156	4, 5	1, 84	M_0 19.8
5240	CDP-6-deoxy-delta-3,4-glucoseen reductase	c(5960164..5960192)	HBIQS	0, 80	4, 248	0, 6	1, 90	
5353-(5354)	Type III chaperone protein ShcA-(hopA1)	6085756..6085788	HBQS	1, 2	71, 172	0, 2	1, 328	CM-FCM 17.5

5616-(0474-0473)	Hypothetical protein-(HopAS1 (interrruption-N)-HopAS1 (interruption-C))	522444..522475	HBS	75, 79	4, 34	2, 6	77, 876	FM* 14.7
5617	conserved Hypothetical gene	939413..939445	HB*S	76, 77	13, 67	3, 45	2, 107	M* 20.2
5618	Type III effector pseudogene hopAT1	c(922925..922953)	HIQS	29, 79	4, 214	6, 6	37, 37	M_1 <6.0
5619-(0901)	PSPTO_5619-(hopAG1)	981177..981209	HBS	6, 6	122, 122	x	0, 0	CM_0 11.7 Y
5620-(1568)	PSPTO_5620-(hopAF1)	1731421..1731453	HBS	75, 77	8, 114	5, 5	4, 4	M_0-CM* 17 Y
5622	PSPTO_5622	1548389..1548420	HBS	2, 5	8, 20265	3, 5	1, 1512	M_0 16.1 Y
5623	PSPTO_5623	5355530..5355562	HBS	3, 5	50, 1939	5, 5	6, 6	M 10.9
5633	conserved protein of unknown function	15821..15849	HBIQS	1, 7	8, 31927	x	0, 0	
A0005	type III effector HopAM1-2	c(6595..6627)	HBS	41, 77	134, 218196	x	x	FM 18.4
A0012	type III effector HopX1	c(16103..16135)	HBS	31, 78	134, 18250	x	x	FM 18.1
A0017-(A0018-A0019)	type III chaperone ShcO1-(hopO1-1-HopT1-1)	19658..19690	HBS	2, 4	134, 61326	x	x	M 16.4
B0003	(identical to PSPTO_5633)							
B0078	hypothetical protein	c(6894..6926)	HBS	32, 79	49, 166407	x	x	M_2 16.0

Table 2. 8 Data for new *hrp* promoters

Operon: PSPTO identifier for gene immediately downstream from promoter.
Function: annotated function for operon-identifying gene.
Coordinate: DC3000 genome coordinate for the region bracketing the –35 and –10 regions of the promoter. "c" designates that the promoter is found on the complementary strand.
Evidence: Experimental evidence for *hrp* promoters from this study.

- H: *hrp* promoter motif found;
- B: Binding activity for HrpL (ChIP-qPCR) observed;
- I: Induction observed in promoter fusion (threshold = 2.4; 2 x negative control);
- S: mRNA 5'-end captured (TSS) within 10 bps from the 3' end of -10 promoter element. No threshold is applied. Absolute values for read counts appear in Table 2.7. TSSs from Filiatrault *et al*. [62] were also taken into consideration.

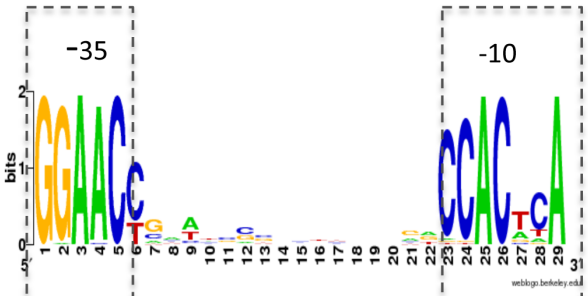
qRT-PCR: Transcript abundance for regions downstream from *hrp* promoters in DC3000 compared to that in a *ΔhrpL* strain. Values in brackets indicate that abundance was measured upstream from coding region. Values are average of one or two biological replicates with three technical replicates and standard deviation.
Published data:

- Mucyn *et al.*: '+' indicates that the gene was classified as differentially expressed in Table 3 and Table S3 by Mucyn *et al.* [130]
- Promoters: symbols indicate whether *hrp* promoters were reported by Chang *et al.* [85] (C), Fouts *et al.* [80] (F), or by Ferreira *et al.* [65] (M).
- Genes: symbols indicate whether genes downstream from *hrp* promoters were tested and/or reported by Chang *et al.* [85] (C), Fouts *et al.* [80] (F), or by Ferreira *et al.* [65] (M). Genes that showed no differential expression, or which were not tested by Ferreira *et al.* [65] are designated (M_o) and (M_1), respectively.

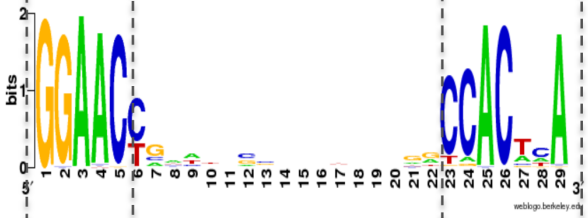
No binding evidence for HrpL or upstream *hrp* promoter motifs were observed in association with 12 of the 14 genes proposed by Mucyn *et al.* [130] as putative novel HrpL regulon members. These are PSPTO_0829 (clpB protein), PSPTO_0851 (hypothetical protein), PSPTO_1371 (effector locus protein), PSPTO_2129 (sensory box histidine kinase/response regulator), PSPTO_2208 (heat shock protein HtpG), PSPTO_3148 (magnesium chelatase subunit ChII), PSPTO_4210 (ATP-dependent protease La), PSPTO_4332 (hypothetical protein), PSPTO_4376 (chaperonin, 60 kDa), PSPTO_4505 (dnaK protein), PSPTO_4716 (hypothetical protein), and PSPTO_4723 (hypothetical protein).

Operon	Function	*hrp* promoter coordinate	Evidence	qRT-PCR	Published data		
					Mucyn *et al.*	Promoters	Genes
371	iaaL indoleacetate-lysine ligase	406210..406238	HBIS	1151.0±177.6	+	FM	FCM_o
871	macrolide efflux protein, putative	939675..939703	HBIS	31.9±3.6	+		M_o
1645	MarR family transcriptional regulator	1802305..1802333	HBIS	(54.6±14.6)*		M	M
1843	aspartate kinase	2012108..2012136	HBIS	2.2±0.2	+		M_o
2130	LuxR family DNA-binding response regulator	c(2304331..2304359)	HBIS	57.5±11.2	+	M	M_o
2691	TerC family membrane protein	2984435..2984463	HBIS	49.9±6.0	+		M
2696	mutT/nudix family protein	(2990249..2990277)	HBIS	29.2±5.0		M	M_o
3331	protease inhibitor Inh	c(3768950..3768978)	HBIS	17.1±5.2			M_o
3481	Hypothetical protein	3929005..3929032	HBS	(15.6±1.6)*			M_1
3721	fabI enoyl-(acyl-carrier-protein) reductase	4199604..4199632	HBIS	(1.4±0.1)*		M	M_o
3948 3949	Intergenic region	c(4457004..4457031)	HBI	(5.0±1.4)*			M_1
4340	insecticidal toxin protein, putative	c(4895201..4895228)	HBIS	(3.1±0.3)*			M_o
4699	non-ribosomal peptide synthetase, terminal	c(5328022..5328050)	HBIS	223.4±57.9		M	M_o
4721	Type III chaperone ShcV	c(5346761..5346788)	HBIS	439.7±164.1	+		M_o
4750	Hypothetical protein	5384480..5384507	HBIS	62.8±17.8			M_o
4955	bifunctional thiosulfate	5616671..5616699	HBIS	13.7±3.6		M	M_o
5053	Hypothetical protein	5751475..5751504	HBIS	(209.7±50.5)*		M	M_o
5240	CDP-6-deoxy-delta-3,4-glucoseen reductase	c(5960164..5960192)	HBIS	8.6±1.0			M_o
5618	Type III effector pseudogene hopAT1	c(922925..922953)	HIS	99.9±7.2		M	M_1
5633	conserved protein of unknown function	15821..15849	HBIS	61.2±9.4			M_1
B0003	(identical to PSPTO_5633)						

A.



B.



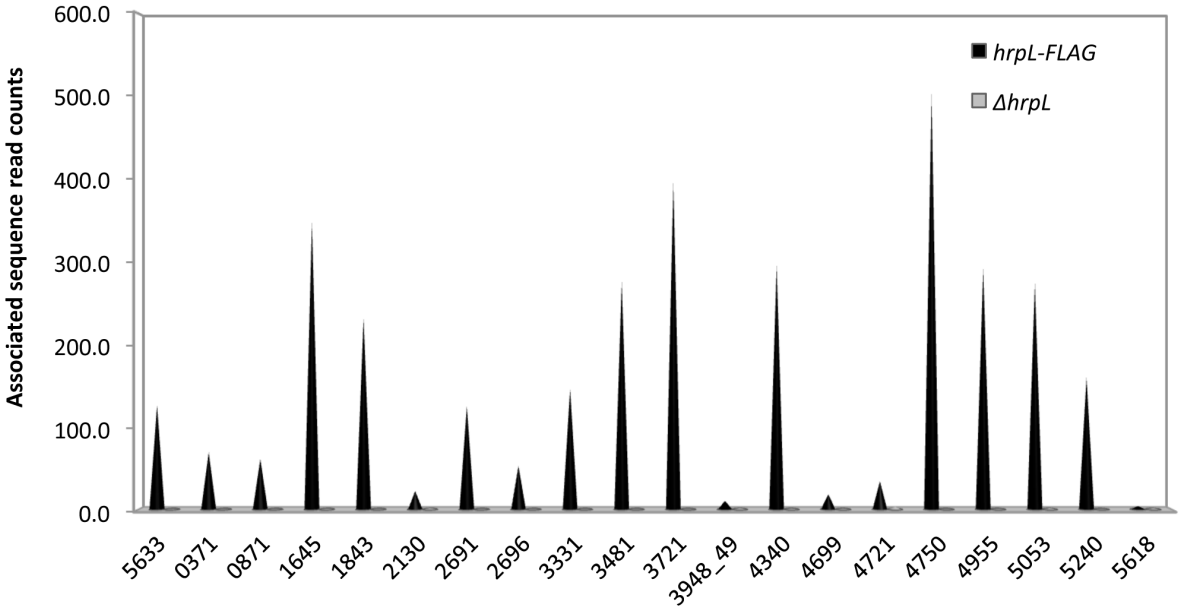
C.

	ID	Motif coordinates 5'	3'	PSPTO
GGAACTTGCCGGCTGGTGCGGCTCCACTAA	P_15835	15821	15849	5633
GGAACCGATCTCGCCATCGAGGTAACTCA	P_406224	406210	406238	0371
GGAACCCATCAGCACAAGCGCACCACGAA	P_939689	939675	939703	0871
GGAACTTGCGCCGCGACCTCAAACCACAAA	P_1802319	1802305	1802333	1645
GCAACCAAACTGAACCATTCGGCCACTCA	P_2012122	2012108	2012136	1843
GGAACTTGAGTGACGGAAACGTTTCACTTAA	P_2304345	c(2304331	2304359)	2130
GGAACTTTCGGTACAGGAGCTACCACAAA	P_2984449	2984435	2984463	2691
GGAACTGTTTTCTGGCAATGGCCCACTTGA	P_2990263	(2990249	2990277)	2696
GGAACTGTTGGTCAAGCGTTTGCCACTTGA	P_3768964	c(3768950	3768978)	3331
GGAACTTGATTCCCAGGATCGG-TGACTTAA	P_3929018	3929005	3929032	3481
GGAACTTGGTCGGCATCCCTGAGCCACATA	P_4199618	4199604	4199632	3721
GGAAGGACATCCTCTTCGTGA-TCACCTG	P_4457017	c(4457003	4457031)	3948_9
GGAACTTCAAGGCTGACGCGG-TAACTCA	P_4895214	c(4895201	4895228)	4340
GGAACCGATAACAAATTTTTGGCCACATA	P_5328036	c(5328022	5328050)	4699
GGAACGTGTCCGCTGTTACGG-CCACTTA	P_5346774	c(5346761	5346788)	4721
GGAACCCATGGAGCTTGTTAG-CCACATA	P_5384493	5384480	5384507	4750*
GGAACCATTGAGCCTGTTCGCGCCACTGA	P_5616685	5616671	5616699	4955
GGATCAAAAATGGGCCTGGATGCCACTTAA	P_5960178	c(5960164	5960192)	5240
GGAACCAAATTGCTGGCTACGCTAACCAA	P_922939	c(922925	922953)	5618
GGAACNNNNNNNNNNNNNNNCCACNNA	consensus			
GGAAACCGGACAGCAGTTGCTGGCCACTCA	P_5751489	5751475	5751504	5053

Figure 2. 3 *hrp* promoter sequence alignment.

(A). Motif logo for annotated *hrp* promoter sequences. (B). Motif logo for all *hrp* promoters including the newly identified set. The –35 region is highly conserved but two cytosines in the –10 region show variability. (C). Alignment of individual motifs sequences. Motif ID is the central position (genome coordinate) for the associated ChIP-Seq zone of enrichment. Genes downstream of *hrp* promoters are identified by PSPTO numbers. (*): Candidate *hrp* promoter is oriented in an antisense direction relative to PSPTO_4750. PSPTO_3948-9: candidate *hrp* promoter is between PSPTO_3948 and PSPTO_3949, which are oriented covergently. Motif logos were created by Weblogo [158]. Sequences were aligned and visualized using SeaView [159].

A.


Associated sequence read counts
600.0
500.0
400.0
300.0
200.0
100.0
0.0
hrpL-FLAG
ΔhrpL
5633
0371
0871
1645
1843
2130
2691
2696
3331
3481
3721
3948_49
4340
4699
4721
4750
4955
5053
5240
5618

B.

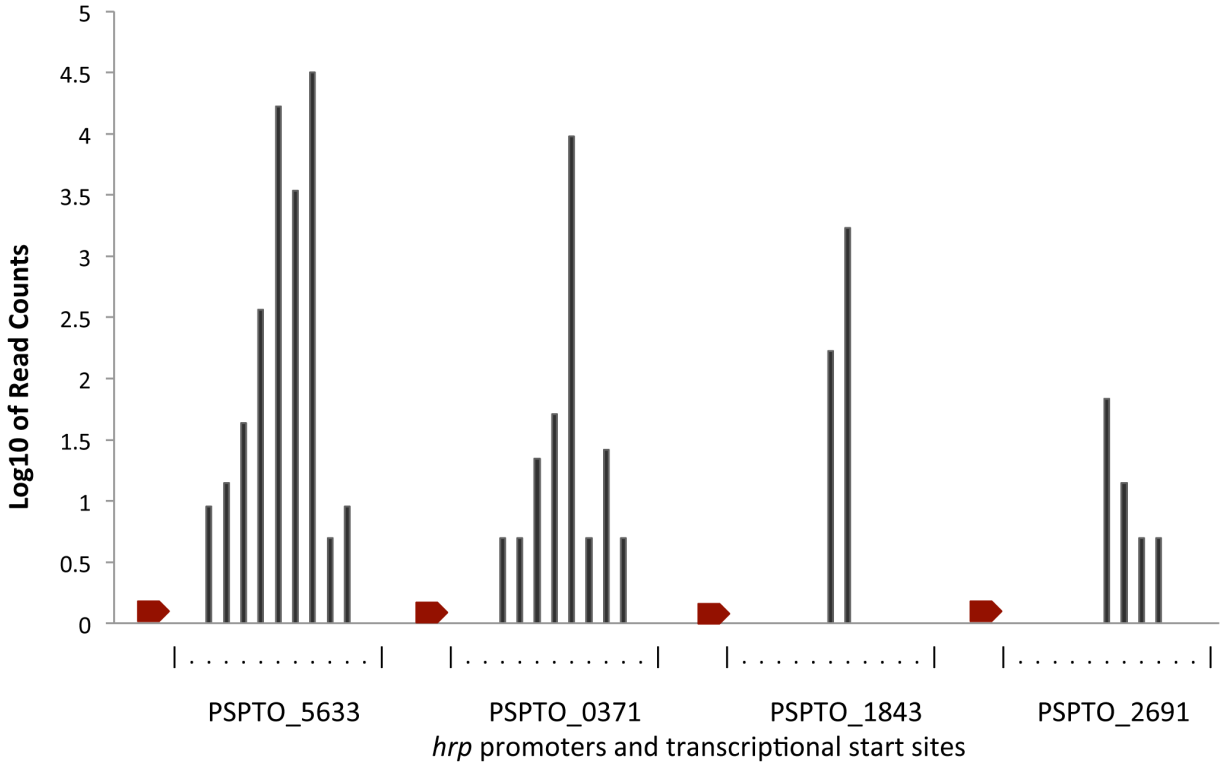

Log10 of Read Counts
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
PSPTO_5633
PSPTO_0371
PSPTO_1843
PSPTO_2691
hrp promoters and transcriptional start sites

Figure 2. 4 Associated sequence read counts at HrpL-binding sites.

(A) The height of each bar corresponds to the number of sequence reads associated with each site enriched by ChIP-Seq, normalized by the number of reads surrounding the peak region (see Materials and Methods). Heights for *hrpL-FLAG* and *ΔhrpL* strains are plotted for comparison. (B) Transcription start site clusters at *hrp* promoters upstream of PSPTO_5633, PSPTO_0371, PSPTO_1843, and PSPTO_2691. Red arrows depict *hrp* promoters. Read counts for captured 5'-ends are shown at positions downstream from each promoter. The mini-scales on the x-axis are positioned so that the leftmost vertical bar corresponds to the 3' nucleotide of the –10 region in Figure 2.3 for the corresponding promoter. Dots indicate individual nucleotide coordinates.

Because of the high resolution afforded by ChIP-exo, we were able to align enriched sequences and identify by inspection a conserved motif that resembles the accepted sequence for the HrpL-responsive promoter, GGAAC(–35)-$N_{16\text{-}17}$-CCACNNA(–10), particularly in the –35 region (Figure 2.3). The region upstream of PSPTO_5053 (ID: P_5751489) appears to have an atypical –35 region (GGAAAC) and is longer than the others by 1 nucleotide. This subtle change at –35, however, can be tolerated by extracytoplasmic function (ECF) sigma factors, of which HrpL is an example [160, 161]. Associated sequence read counts at HrpL-binding sites were also computed to represent magnitude of enrichment (Figure 2.4A). All enrichment values from the *hrpL-FLAG* strain (ranging from 4.7 for P_922939 to 502.0 for P_5384493) are typically much larger than the corresponding values from the *Δ hrpL* strain (0.0 for most cases; 4.0 for P_5346774). The resemblance of the candidate promoters to the canonical *hrp* promoter consensus sequence, together with the evidence that HrpL binds at their genomic locations suggests that they are genuine HrpL-responsive promoters.

Genome-wide identification of TSSs in RNA-Seq

Active promoters should be associated with nearby transcription start sites (TSSs). Using high-throughput sequencing methods, we mapped TSSs by modifying a stranded RNA-Seq protocol to enrich for primary mRNA transcripts (i.e., those bearing a 5'-triphosphate group) [62]. A unique 18 nucleotide tag was ligated to the 5' end primary transcripts so that they could be readily identified (see Materials and Methods). The resulting RNA-Seq data contain both whole transcriptome and TSS information (Table 2.6).

To identify conserved motifs associated with TSS, the 1500 tagged ends with the most abundant reads were selected for analysis. We extracted 50 nucleotides directly upstream from each captured 5'-end, resulting in 1451 sequences derived from the *hrpL-FLAG* sample and 1472 sequences from the *ΔhrpL* sample (overlapping sequences within a sample were merged) and used the sequences as input to MEME [154]. Three motifs with structures resembling canonical promoters were detected in the *hrpL-FLAG* sequences (Figure 2.1). These include likely RpoD (motif 1, 5'-ttG-N_{20}- TANaaT-3') [134] and RpoF (motif 2, 5'-TaAaG-N_{15}-GcCGAta-3') [162] promoters as well as the putative *hrp* promoter (motif 3, 5'-GgAACc –$N_{16\text{-}17}$-CCAN-3') [65]. We also recovered a weak motif resembling the RpoN promoter (motif 4, 5'-TgG-N_{10}-TGC-3') [163]. Two motifs were recovered from the *ΔhrpL* samples, motif 1 (RpoD) and motif 2 (RpoF). The putative RpoF promoters derived in both cases are upstream of genes likely to be regulated by this sigma factor, such as *fliC* (PSPTO_1949), *fliD* (PSPTO_1951), *flgK* (PSPTO_1944), *cheY-2* (PSPTO_1980) and *cheA-2* (PSPTO_1982) (see supplemental data). As expected, the HrpL motif was identified in the *hrpL-FLAG* but not the *ΔhrpL* samples.

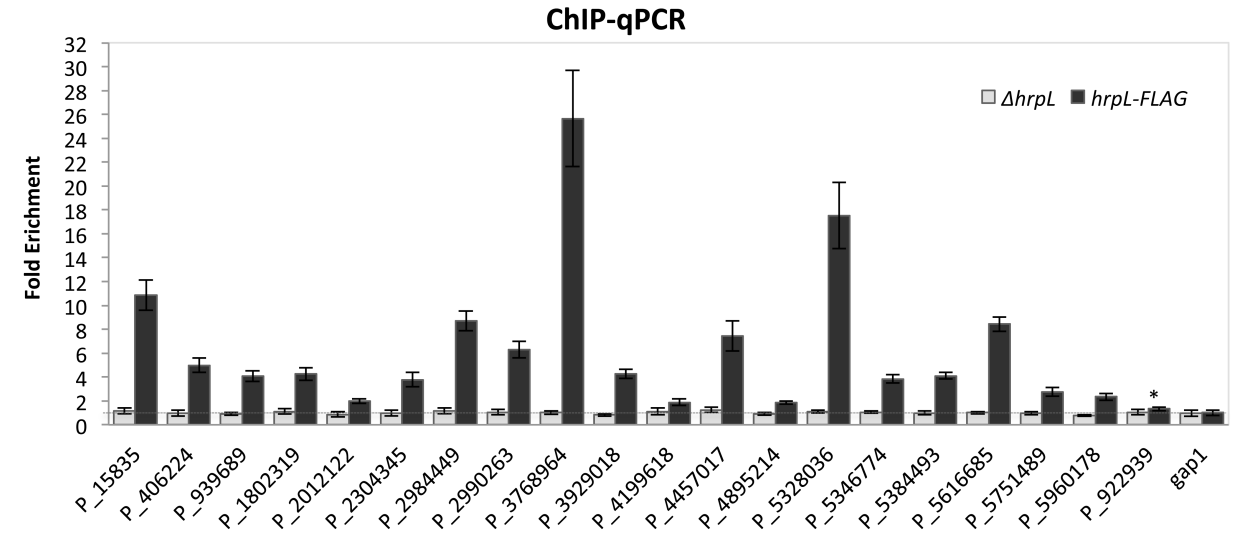
The overall number of detected promoter motifs contrasts with results reported by Filiatrault *et al.* [62], who identified nine distinct motifs upstream from captured 5'-ends (including the *hrp* promoter) using the similar MEME parameters but different DC3000 culture conditions and sample preparation methods. However, in agreement with that report, we also observed that captured ends tend to occur in tightly spaced clusters (see example in Figure 2.4B). TSS clustering has been noted in other bacteria [164, 165] and in DC3000 for both PvdS and HrpL-associated promoters [62]. The most abundant TSS signal, as well as the signal closest to each promoter candidate, is shown in Table 2.7. In bacteria,

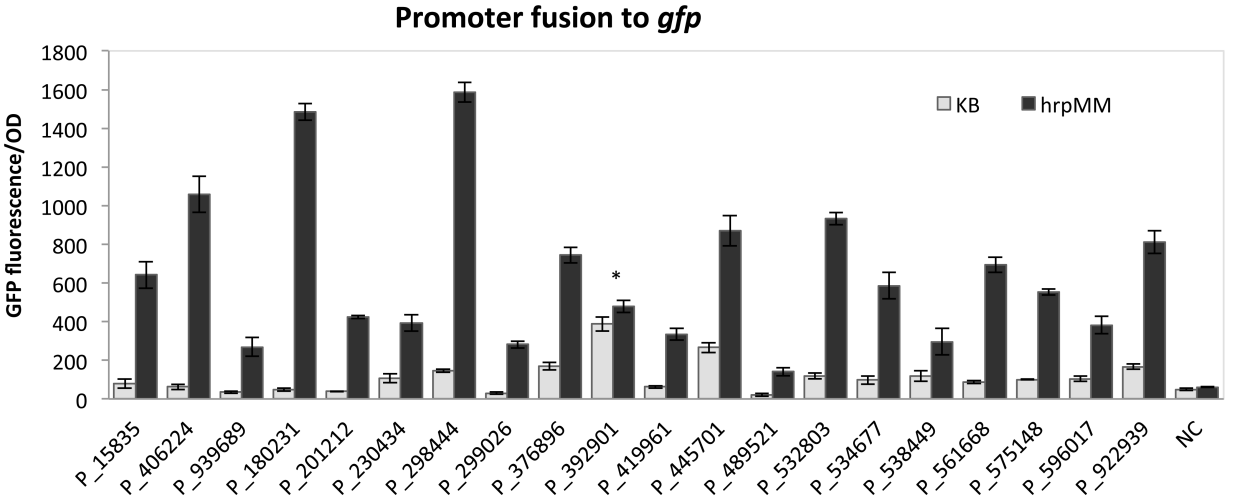
transcription begins 10 bps or less downstream from -10 promoter elements [56]. In four cases the distance between a *hrp* promoter candidate and its closest captured 5'-end exceeds this limit (candidates upstream of PSPTO_1022, PSPTO_A0005, PSPTO_A0012, and PSPTO_B0078). Although these promoters may simply be inactive, it is also possible that the TSSs are undetectable because they are rapidly degraded in the cell or during sample processing. Previous experimental evidence suggests that all four of these *hrp* promoters are genuine [65, 80, 85]. Another *hrp* promoter candidate, between PSPTO_3948 and PSPTO_3949, is approximately 30 bps distant from the closest TSS. Validation tests suggest that this promoter supports HrpL-dependent transcription (below). In summary, all 20 candidate *hrp* promoters are associated with regions enriched by ChIP-Seq and a plausible *hrp* promoter motif; most are positioned close to captured 5'-ends (between 0-7 bps). It is therefore likely that these represent *bona fide* HrpL-dependent promoters.

Multiple methods validate candidate HrpL-dependent promoters

The twenty *hrp* promoters discussed above were analyzed further using ChIP-qPCR, reporter fusions and qRT-PCR to confirm HrpL binding, promoter function and HrpL-dependent transcription.

Cells for ChIP-qPCR were cultured and collected independently from those used for ChIP-Seq. Analysis of the DNA isolated by immunoprecipitation confirmed enrichment at 19 out of 20 targets (Figure 2.5A). The exception was P_922939, located upstream of PSPTO_5618 (pseudogene for HopAT1). This promoter also has a relatively weak signal in the ChIP-Seq experiment. However, PSPTO_5618 was transcribed in a HrpL-dependent manner in our other tests (described below).


ChIP-qPCR
Fold Erichment
ΔhrpL
hrpL-FLAG
P_15835
P_406224
P_939689
P_1802319
P_2012122
P_2304345
P_2984449
P_2990263
P_3768964
P_3929018
P_4199618
P_4457017
P_4895214
P_5328036
P_5346774
P_5384493
P_5616685
P_5751489
P_5960178
P_922939
gap1


Promoter fusion to gfp
GFP fluorescence/OD
KB
hrpMM
P_15835
P_406224
P_939689
P_180231
P_201212
P_230434
P_298444
P_299026
P_376896
P_392901
P_419961
P_445701
P_489521
P_532803
P_534677
P_538449
P_561668
P_575148
P_596017
P_922939
NC

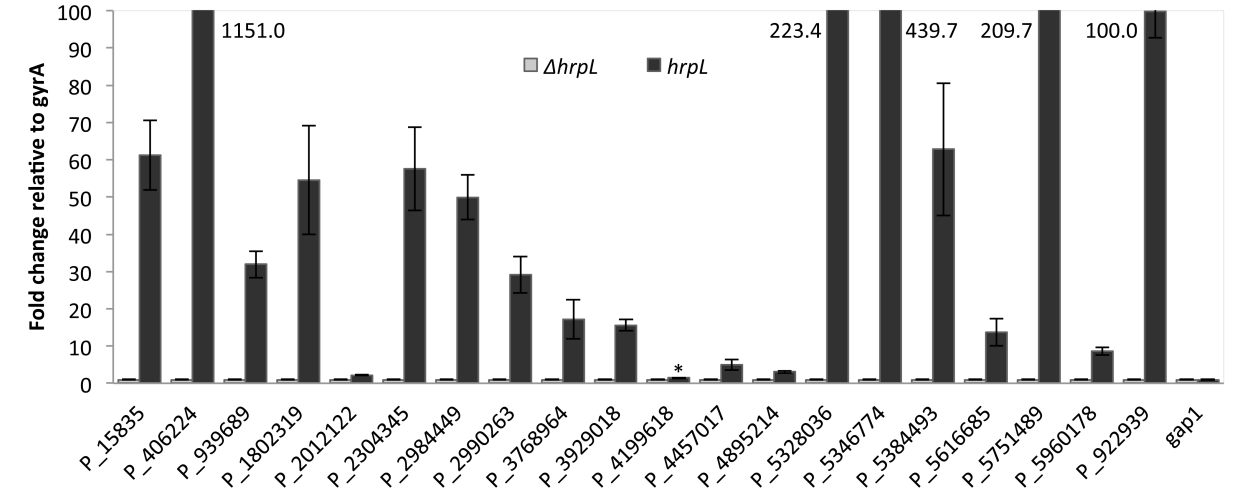

Relative Induction
Fold change relative to gyrA
1151.0
223.4
439.7
209.7
100.0
ΔhrpL
hrpL
P_15835
P_406224
P_939689
P_1802319
P_2012122
P_2304345
P_2984449
P_2990263
P_3768964
P_3929018
P_4199618
P_4457017
P_4895214
P_5328036
P_5346774
P_5384493
P_5616685
P_5751489
P_5960178
P_922939
gap1

Figure 2. 5 Validation of new *hrp* promoters

(A). ChIP_qPCR experiments to test enrichment of DNA fragments at putative HrpL binding sites. Values for each gene were normalized to results for *gyrA* (DNA gyrase subunit A). *gap-1* (glyceraldehyde 3-phosphate dehydrogenase, type I), not predicted to be HrpL-regulated, was used as a negative control. All fold changes above the expression value for *gyrA* are classified as enriched (above the horizontal line). (B). Induction of cloned *hrp* promoter-*gfp* fusions. Induction was measured by relative fluorescence normalized by OD_{600} (GFP fluorescence/OD) in *hrp*-inducing and *hrp*-repressing conditions. The *hrp* promoter::*gfp* fusion constructs were expressed in the DC3000 Δ*pvsA* siderophore mutant. The promoter trap vector without a promoter insert was used as a negative control (NC). GFP was measured using a Synergy 2 plate reader (Biotech) with excitation from 475 to 495 nm and emission from 506 to 526 nm. OD was measured at 600 nm using the same plate reader. A kinetics reading procedure was used, and a single data point at 5 hours was plotted for all strains, which is the time at which they show a peak value. (C). qRT-PCR analysis showing HrpL-dependent differential expression of transcripts downstream from *hrp* promoters in WT DC3000 and *Δ hrpL* strains. The relative fold change was measured after 1.5 hours on MG supplemented with iron (50 μM final concentration) normalized to *gyrA.* For determination of the relative expression, expression of each gene in the ΔhrpL mutant was set to 1. Expression of each gene in the WT strain was then normalized to the corresponding gene in the ΔhrpL mutant. All data points are the averages of 3 replicates with standard deviations.

A plasmid-based reporter trap assay was designed to test promoter activity at defined chromosomal regions and to isolate promoter activity from potentially confounding effects such as read-through from transcription initiated further upstream. Genomic fragments of 100 to 200 nucleotides containing candidate promoters were used to construct transcriptional fusions with a green fluorescent protein (GFP) reporter. Although the basal expression of each promoter was different on KB, all twenty regions showed strong induction in *hrp*-inducing medium compared to KB (Figure 2.5B). P_3929018 (PSPTO_3481) exhibits robust expression on both rich medium and *hrp*-inducing medium, suggesting that the cloned region contains a promoter that is constitutively expressed. However, qRT-PCR shows that transcription of PSPTO_3481 is 15-fold higher in DC3000 vs. Δ*hrpL* (below). A *hrp* promoter motif at this location was identified in an earlier global transcriptional map [62]. It is possible that this region contains HrpL-dependent and independent promoters.

A qRT-PCR assay was used to assess HrpL-dependent transcription in regions downstream from candidate *hrp* promoters in their native genomic context. Relative transcript abundance for candidates was computed by comparing relative transcript levels at 1.5 hours after shifting bacterial cells from KB to MG (iron supplemented) medium, normalized to mRNA levels for *gyrA*, a housekeeping gene (Figure 2.5C). Most previously reported *hrp* promoters were associated with strong induction in this experiment (Figure 2.6). When the relative induction of transcripts in the Δ*hrpL* mutant is compared to that of WT DC3000 in *hrp*-inducing medium, genes downstream from 18 of the new promoter candidates show a strong HrpL-dependent expression. P_406224 (upstream of

PSPTO_0371) showed the largest induction in this experiment (> 1000x), with four others showing levels ≥ 100x. P_2012122 (upstream of PSPTO_1843) shows only small differences in mRNA levels between the two backgrounds. P_4199618, upstream of PSPTO_3721, showed no significant induction using qRT-PCR but did exhibit HrpL-dependent behavior in the reporter trap assay. This promoter may depend on other factors for transcription or may ordinarily function at a very low level of activity.

P_5384493 was verified in all experiments. This *hrp* promoter candidate is in an intragenic region downstream of PSPTO_4750 (a hypothetical protein) but is oriented in the antisense direction. As shown in Figure 2.7, the captured 5' TSS in close proximity to this promoter is consistent with antisense transcription. Although our qRT-PCR protocol cannot distinguish between transcripts arising from complementary regions, the candidate promoter cloned into our GFP fusion construct showed 2.5 fold induction in *hrp*-inducing compared to KB rich medium.

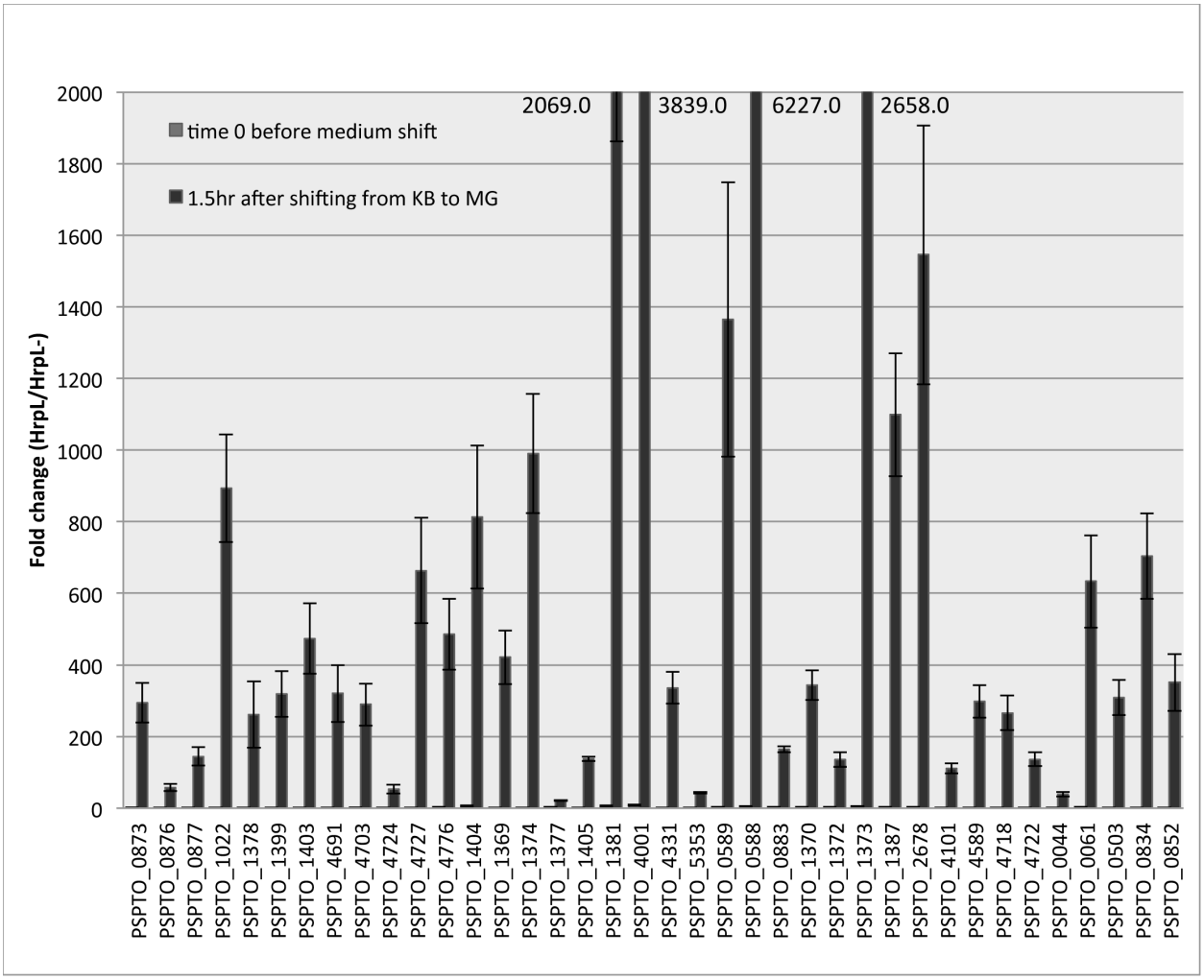


Figure 2. 6 qRT-PCR analysis showing HrpL-dependent transcription downstream from 38 known HrpL regulon members.

Relative transcript change was compared between DC3000 and *ΔhrpL* strains. Relative induction of each gene was normalized to the housekeeping gene *gap1.* No transcription induction was observed in KB, while significant induction was seen after medium shift to MG supplemented with iron (50 μM final concentration) after 1.5hr. Values are averages of three replicates with standard deviations.

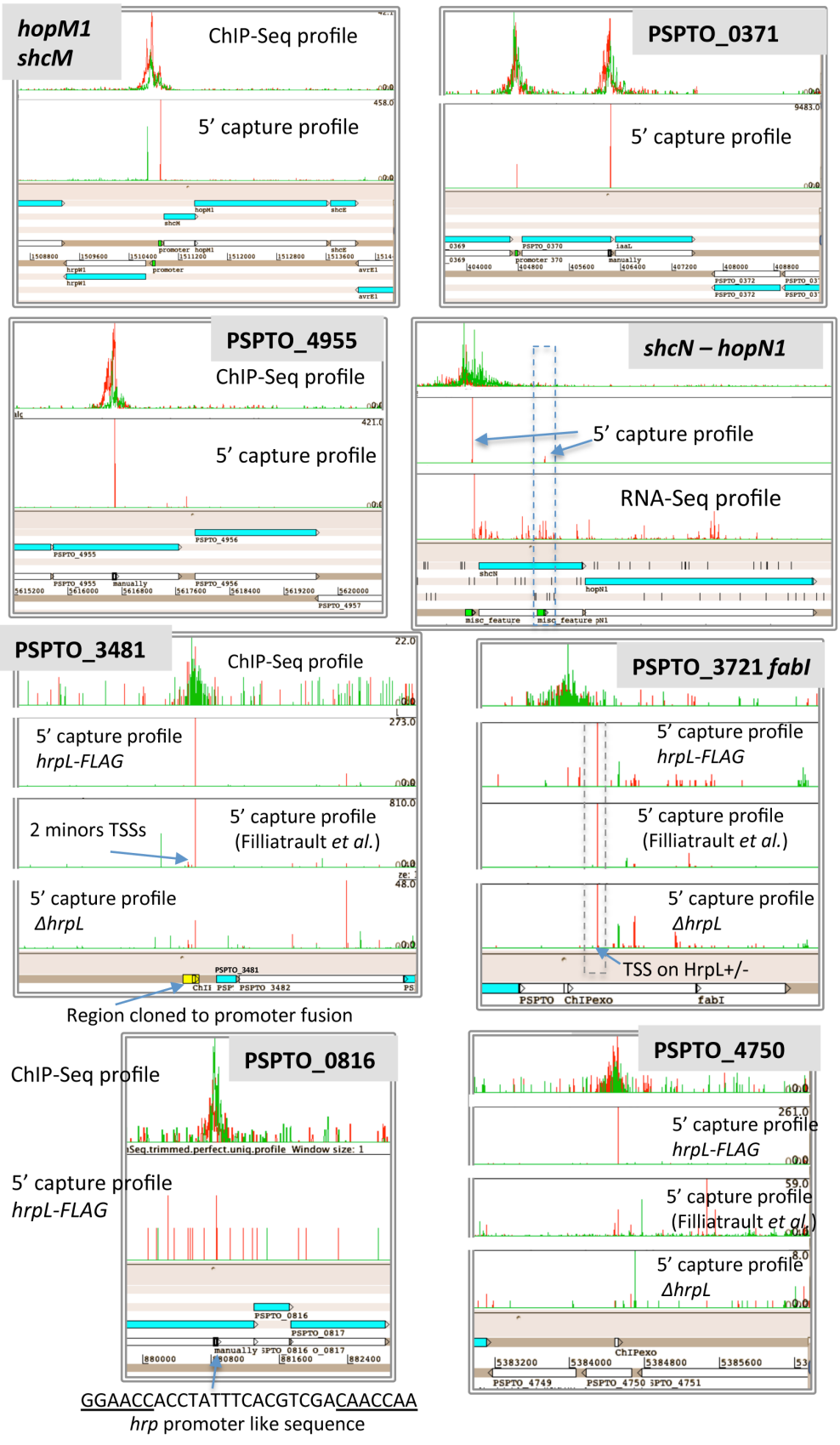

hopM1
shcM
ChIP-Seq profile
5’ capture profile
PSPTO_0371
5’ capture profile
PSPTO_4955
ChIP-Seq profile
5’ capture profile
shcN – hopN1
5’ capture profile
RNA-Seq profile
PSPTO_3481
ChIP-Seq profile
5’ capture profile
hrpL-FLAG
2 minors TSSs
5’ capture profile
(Filliatrault et al.)
5’ capture profile
ΔhrpL
Region cloned to promoter fusion
PSPTO_3721 fabI
5’ capture profile
hrpL-FLAG
5’ capture profile
(Filliatrault et al.)
5’ capture profile
ΔhrpL
TSS on HrpL+/-
ChIP-Seq profile
PSPTO_0816
5’ capture profile
hrpL-FLAG
GGAACCACCTATTTCACGTCGACAACCAA
hrp promoter like sequence
PSPTO_4750
5’ capture profile
hrpL-FLAG
5’ capture profile
(Filliatrault et al.)
5’ capture profile
ΔhrpL

Figure 2. 7 ChIP-Seq and RNA-Seq data for selected HrpL regulon members.

The red line represents mapped reads corresponding to the positive strand and the green line shows reads mapped to the negative strand. Genome annotation is shown below profiles. The sequence containing the *hrp* promoter motif is represented by green boxes. The 5' capture profile from Filliatrault *et al.* [62] is included in some panels for comparison.

Results compared with a computational inventory of *hrp* promoters

Sequence pattern matching has been used extensively to inventory the HrpL regulon in DC3000 [65, 79, 86]. To help determine whether the procedures described above identified all DC3000 *hrp* promoters, we scanned the DC3000 genome using a hidden Markov model trained using previously annotated and new *hrp* promoter sequences (Table 2.3) and compared the results to those obtained by ChIP-Seq and RNA-Seq. Although the scan matched all annotated and new candidate *hrp* promoters identified in this study, the model did not match any other region in the genome that showed enrichment in the ChIP-Seq experiment (E-value cut-off = 0.001, 245 promoter candidates in total). As a further precaution, the E-value threshold for a match was reset to 0.01, an even more relaxed level that predicts 424 *hrp* promoters in the DC3000 genome. Among matches that are unassociated with confirmed promoters, five are in regions where ChIP-Seq shows weak enrichment (Table 2.5). These genes are not associated with detectable TSSs within an appropriate distance. The *hrp* promoter-like sequence upstream of PSPTO_0816 (Type IV pilus biogenesis protein) was examined more closely. This candidate shows no significant enrichment using ChIP-qPCR, and does not demonstrate HrpL-dependent transcription using qRT-PCR (data not shown). However, a transcriptional promoter fusion involving this region exhibits a 10.9 fold induction in *hrp*-inducing medium (Table 2.5). The candidate *hrp* promoter is similar to the canonical *hrp* promoter within the –35 region, but varies at two bases within the –10 region (CAACCAA instead of CCACNNA; Table 2.5). While it is possible that the induction is due to this promoter candidate rather than some other cryptic promoter, the data are equivocal and the candidate has not been included in Table 2.8. We

suspect that other candidates identified using extremely relaxed criteria will be similarly difficult to classify with confidence.

The new members of the DC3000 HrpL regulon are largely unrelated to virulence

The annotated functions of the genes associated with the new *hrp* promoters do not appear to involve the T3SS machinery or add to the effector repertoire, with the possible exception of PSPTO_5633 (see below). In order to determine whether the new genes contribute to pathogenicity, we constructed deletion mutants (in WT DC3000 and *ΔhopQ1-1* backgrounds) for seven candidates whose annotated functions were suggestive of plant association and examined them for an altered virulence phenotype in *N. benthamiana* (PSPTO_5633, PSPTO_0371, PSPTO_2691, PSPTO_2696, PSPTO_3331, PSPTO_5240 and PSPTO_2130). No phenotypic differences were observed for *in planta* growth, virulence, or HR for any strain (data not shown). The result for PSPTO_2130 is consistent with another analysis reported for this gene [130]. Although it is not uncommon for effector mutants to fail to exhibit a phenotype (due to functional redundancy [89, 90, 166]), most of these genes are unlikely to be effectors (see Discussion). It is possible that some of the remaining 13 candidates will demonstrate phenotypes if mutated and tested.

PSPTO_5633 appears to be a weak Type III secretion effector

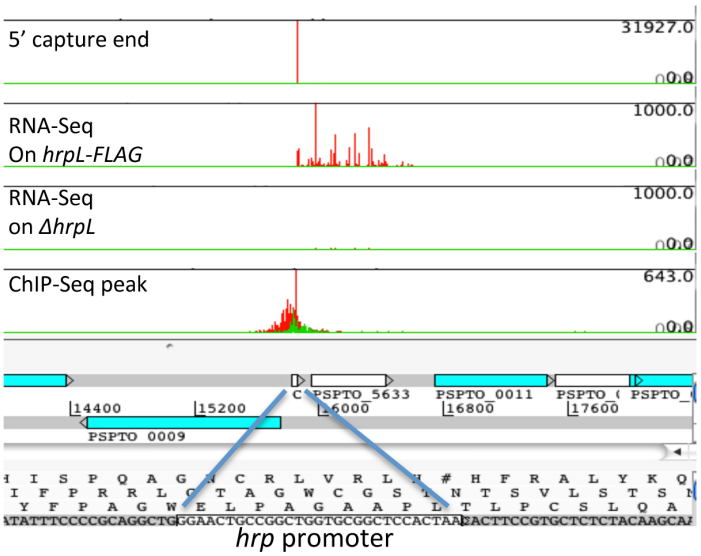
Although PSPTO_5633 has no annotated function, it shares high sequence similarity with hypothetical proteins in some pathogenic bacterial species (such as *P. syringae* pv. *maculicola*, *Erwinia tracheiphila*, *Citrobacter rodentium, Burkholderia phymatum, Xanthomonas campestris* and *Yersinia mollaretii*), which is frequently the case for an

effector. However, the TEREE (Type III Effector Relative Entropy Evaluation) score for this gene is –2 [167], which is outside of the range exhibited by most effectors (–13 to –47). Nonetheless we investigated PSPTO_5633 as a potential effector. Figure 2.8A shows PSPTO_5633 and its neighboring *hrp* promoter motif, aligned with the ChIP-Seq, RNA-Seq and 5'-capture profiles in this genomic region.

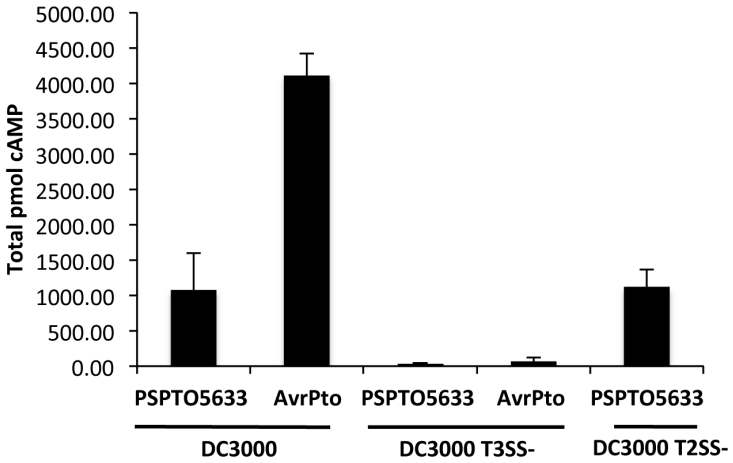
As effectors need to be transported into the plant cytoplasm to interact with plant defenses, we tested PSPTO_5633 for its ability to translocate into plant cells. Figure 2.8B shows that PSPTO_5633 enters plant cells in a T3SS dependent manner. PSPTO_5633 translocation is weak when compared to that for AvrPto but is significantly above background, a result supported by multiple experimental replicates. Based on these results, PSPTO_5633 has been assigned the name HopBM1 to recognize its role as an effector.

A closer look at the N-terminal region of PSPTO_5633 surprisingly suggests that this protein may be secreted through the Sec pathway (see Figure 2.8C). To test this possibility, translocation of PSPTO_5633 was assayed in a Δ*gspD* background (T2SS–). The level of cAMP observed using a T2SS–strain was indistinguishable from that seen with DC3000 (Figure 2.8B), suggesting that PSPTO_5633 does not enter plant cells using this pathway.

A.


5’ capture end
31927.0
0.0
RNA-Seq
On hrpL-FLAG
1000.0
0.0
RNA-Seq
on ΔhrpL
1000.0
0.0
ChIP-Seq peak
643.0
0.0
14400
15200
PSPTO_5633
16000
PSPTO_0011
16800
17600
PSPTO_0009
hrp promoter

B.


Total pmol cAMP
5000.00
4500.00
4000.00
3500.00
3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00
PSPTO5633
AvrPto
PSPTO5633
AvrPto
PSPTO5633
DC3000
DC3000 T3SS-
DC3000 T2SS-

C.

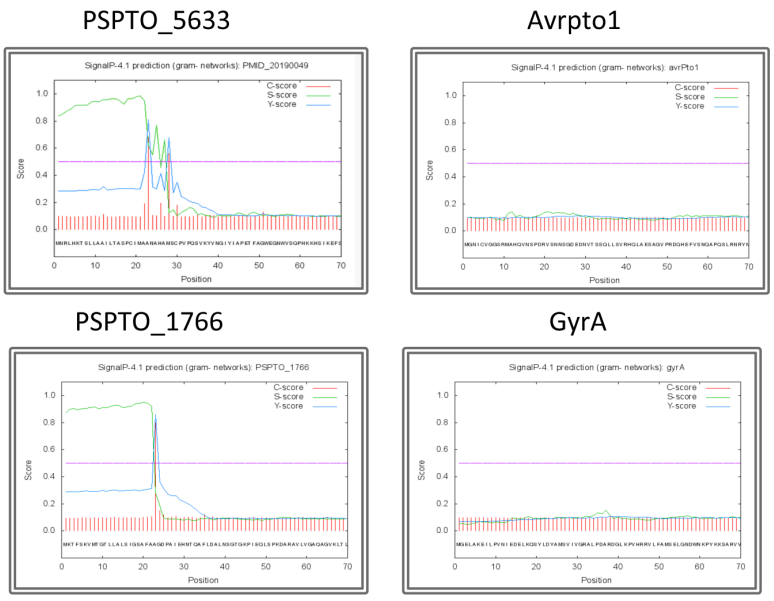

PSPTO_5633
Avrpto1
PSPTO_1766
GyrA
C-score
S-score
Y-score
Score
Position

Figure 2. 8 Summary of data for PSPTO_5633.

(A). ChIP-Seq, RNA-Seq and promoter motif at PSPTO_5633 locus. The transcription start site mapped by 5' capture in RNA-Seq and its location relative to the predicted motif are consistent with the presence of a genuine *hrp* promoter. The profiles, along with genome annotation, are shown using Artemis. Red and green traces correspond to sequence read counts on the positive and negative strands, respectively. The sequence containing the *hrp* promoter motif is enclosed in a box. (B) Evidence that PSPTO_5633 is translocated through the DC3000 T3SS. *N. benthamiana* leaves were infiltrated with $5x10^7$ CFU/ml of the indicated DC3000 strains carrying plasmids in which PSPTO_5633 was fused to the Cya translocation reporter, or an AvrPto-Cya control. Total cAMP produced as a result of Cya activity in leaf extracts 6 hours after infiltration is shown for all the strains. PSPTO_5633 is translocated into leaf cells from wild-type DC3000 ($T3SS^+$) and from a DC3000Δ*gspD* (T2SS- mutant. No translocation was observed in the DC3000Δ*hrcQ-U* ($T3SS^-$ mutant) background. The data represent the average cAMP (pmol) with standard deviations computed using data from 3 plants. The experiment was repeated 3–5 times for all strains except for PSPTO_5633(DC3000 T2SS–), which was repeated twice. (C) SignalP analysis showing C, S and Y scores for each position in the sequence of PSPTO_5633, where C-score is the raw cleavage site score, S-score is the signal peptide score and Y-score is the combined cleavage site score. Similar analyses for avrPto1 (a T3SS-translocated effector), PSPTO_1766 (lipase, generally known to target the Sec pathway), and a housekeeping gene (gyrase, generally known to function inside bacterial cells) are shown for comparison.

Comparative genome analysis of the HrpL regulon

Although a sigma factor regulon can be described as a tightly integrated collection of genes, the composition of the HrpL regulon across species is not rigid. While the core effector delivery system is highly conserved, the effector proteins delivered by it vary considerably [168] in other *P. syringae* pathovars. In order to determine whether the new HrpL regulon members show similar patterns of conservation, we conducted an *in silico* analysis in which we examined genome sequences from 121 members of the *P. syringae* group, most of which are plant pathogens. The group (taxid 136849) is defined by NCBI in their taxonomy database (http://www.ncbi.nlm.nih.gov/taxonomy).

We identified orthologs to the DC3000 genes immediately downstream from the 73 *hrp* promoters and then examined the DNA sequences upstream from them for patterns matching the *hrp* promoter (Figure 2.9). As expected, nearly all of the *P. syringae* group genomes contain HrpL orthologs. *hrp* promoter motifs are observed upstream of most other regulon orthologs (bright red squares) including those corresponding to the new regulon members (shown with names on a green background). Genes for core T3SS functions are widely shared across the group, whereas orthologs for the DC3000 effector genes are not as conserved, reflecting effector diversity. About half of the new regulon members are as conserved as the core T3SS genes. Others (such as PSPTO_5633 and PSPTO_3481) exhibit a different pattern and are found only in DC3000 and a few other *P. syringae* genomes (Figure 2.10). This sparse distribution resembles that found for effectors such as PSPTO_4691 (HopAD1) and PSPTO_4703 (HopAQ1). Note that ortholog absence should be interpreted cautiously since it can reflect errors in genome assembly or

shortcomings in the methods used to identify reciprocal best BLASTP matches, as well as the actual absence of an ortholog in a genome.

A substantially different result is obtained when the analysis is extended to all *Pseudomonadales* genome sequences (1060 organisms total), most of which are neither plant pathogens nor contain HrpL orthologs (Table 2.4). Orthologs for core regulon components are rarely detected outside the *P. syringae* group. In contrast, several new regulon members are widely represented (Figures 2.10 and 2.11). Examples are PSPTO_1843 (aspartate kinase), PSPTO_3721 (enoyl-[acyl-carrier-protein] reductase) and PSPTO_4955 (thiosulfate sulfurtransferase / phosphatidylserine decarboxylase) whose orthologs appear in almost all sequenced strains tested.

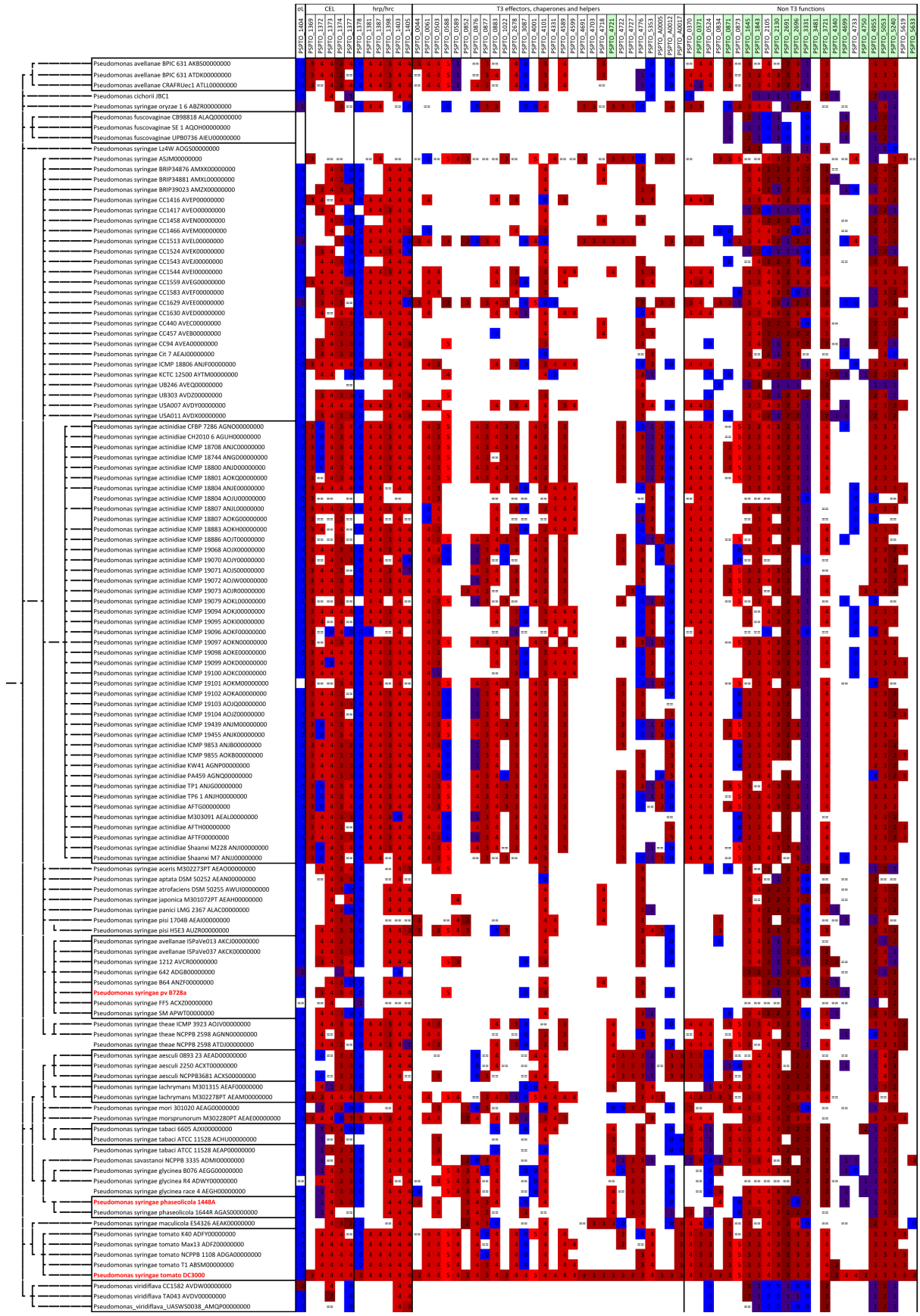

oL
CEL
hrp/hrc
T3 effectors, chaperones and helpers
Non T3 functions
PSPTO_1404
PSPTO_1369
PSPTO_1372
PSPTO_1373
PSPTO_1374
PSPTO_1377
PSPTO_1378
PSPTO_1381
PSPTO_1387
PSPTO_1398
PSPTO_1402
PSPTO_1403
PSPTO_1405
PSPTO_0044
PSPTO_0061
PSPTO_0603
PSPTO_0688
PSPTO_0689
PSPTO_0852
PSPTO_0876
PSPTO_0877
PSPTO_0883
PSPTO_1022
PSPTO_2678
PSPTO_3087
PSPTO_4001
PSPTO_4101
PSPTO_4431
PSPTO_4589
PSPTO_4599
PSPTO_4691
PSPTO_4709
PSPTO_4718
PSPTO_4721
PSPTO_4722
PSPTO_4727
PSPTO_4776
PSPTO_5353
PSPTO_A0005
PSPTO_A0012
PSPTO_A0017
PSPTO_0070
PSPTO_0071
PSPTO_0824
PSPTO_0834
PSPTO_0871
PSPTO_0873
PSPTO_1645
PSPTO_1843
PSPTO_2105
PSPTO_2130
PSPTO_2691
PSPTO_2696
PSPTO_3331
PSPTO_3491
PSPTO_3721
PSPTO_4340
PSPTO_4699
PSPTO_4733
PSPTO_4750
PSPTO_4955
PSPTO_5053
PSPTO_5240
PSPTO_5619
PSPTO_5813
Pseudomonas avellanae BPIC 631 AKBS00000000
Pseudomonas avellanae BPIC 631 ATDK00000000
Pseudomonas avellanae CRAFRUec1 ATLL00000000
Pseudomonas cichorii JBC1
Pseudomonas syringae oryzae 1 6 ABZR00000000
Pseudomonas fuscovaginae CB98818 ALAQ00000000
Pseudomonas fuscovaginae SE 1 AQOH00000000
Pseudomonas fuscovaginae UPB0736 AIEU00000000
Pseudomonas syringae Lz4W AOGS00000000
Pseudomonas syringae ASJM00000000
Pseudomonas syringae BRIP34876 AMXK00000000
Pseudomonas syringae BRIP34881 AMXL00000000
Pseudomonas syringae BRIP39023 AMZX00000000
Pseudomonas syringae CC1416 AVEP00000000
Pseudomonas syringae CC1417 AVEO00000000
Pseudomonas syringae CC1458 AVEN00000000
Pseudomonas syringae CC1466 AVEM00000000
Pseudomonas syringae CC1513 AVEL00000000
Pseudomonas syringae CC1524 AVEK00000000
Pseudomonas syringae CC1543 AVEJ00000000
Pseudomonas syringae CC1544 AVEI00000000
Pseudomonas syringae CC1559 AVEG00000000
Pseudomonas syringae CC1583 AVEF00000000
Pseudomonas syringae CC1629 AVEE00000000
Pseudomonas syringae CC1630 AVED00000000
Pseudomonas syringae CC440 AVEC00000000
Pseudomonas syringae CC457 AVEB00000000
Pseudomonas syringae CC94 AVEA00000000
Pseudomonas syringae Cit 7 AEAJ00000000
Pseudomonas syringae ICMP 18806 ANJF00000000
Pseudomonas syringae KCTC 12500 AYTM00000000
Pseudomonas syringae UB246 AVEQ00000000
Pseudomonas syringae UB303 AVDZ00000000
Pseudomonas syringae USA007 AVDY00000000
Pseudomonas syringae USA011 AVDX00000000
Pseudomonas syringae actinidiae CFBP 7286 AGNO00000000
Pseudomonas syringae actinidiae CH2010 6 AGUH00000000
Pseudomonas syringae actinidiae ICMP 18708 ANJC00000000
Pseudomonas syringae actinidiae ICMP 18744 ANGD00000000
Pseudomonas syringae actinidiae ICMP 18800 ANJD00000000
Pseudomonas syringae actinidiae ICMP 18801 AOKQ00000000
Pseudomonas syringae actinidiae ICMP 18804 ANJE00000000
Pseudomonas syringae actinidiae ICMP 18804 AOJU00000000
Pseudomonas syringae actinidiae ICMP 18807 ANJL00000000
Pseudomonas syringae actinidiae ICMP 18807 AOKG00000000
Pseudomonas syringae actinidiae ICMP 18883 AOKH00000000
Pseudomonas syringae actinidiae ICMP 18886 AOJT00000000
Pseudomonas syringae actinidiae ICMP 19068 AOJX00000000
Pseudomonas syringae actinidiae ICMP 19070 AOJY00000000
Pseudomonas syringae actinidiae ICMP 19071 AOJS00000000
Pseudomonas syringae actinidiae ICMP 19072 AOJW00000000
Pseudomonas syringae actinidiae ICMP 19073 AOJR00000000
Pseudomonas syringae actinidiae ICMP 19079 AOKL00000000
Pseudomonas syringae actinidiae ICMP 19094 AOKJ00000000
Pseudomonas syringae actinidiae ICMP 19095 AOKI00000000
Pseudomonas syringae actinidiae ICMP 19096 AOKF00000000
Pseudomonas syringae actinidiae ICMP 19097 AOKN00000000
Pseudomonas syringae actinidiae ICMP 19098 AOKE00000000
Pseudomonas syringae actinidiae ICMP 19099 AOKD00000000
Pseudomonas syringae actinidiae ICMP 19100 AOKC00000000
Pseudomonas syringae actinidiae ICMP 19101 AOKM00000000
Pseudomonas syringae actinidiae ICMP 19102 AOKA00000000
Pseudomonas syringae actinidiae ICMP 19103 AOJQ00000000
Pseudomonas syringae actinidiae ICMP 19104 AOJZ00000000
Pseudomonas syringae actinidiae ICMP 19439 ANJM00000000
Pseudomonas syringae actinidiae ICMP 19455 ANJK00000000
Pseudomonas syringae actinidiae ICMP 9853 ANJB00000000
Pseudomonas syringae actinidiae ICMP 9855 AOKB00000000
Pseudomonas syringae actinidiae KW41 AGNP00000000
Pseudomonas syringae actinidiae PA459 AGNQ00000000
Pseudomonas syringae actinidiae TP1 ANJG00000000
Pseudomonas syringae actinidiae TP6 1 ANJH00000000
Pseudomonas syringae actinidiae AFTG00000000
Pseudomonas syringae actinidiae M303091 AEAL00000000
Pseudomonas syringae actinidiae AFTH00000000
Pseudomonas syringae actinidiae AFTF00000000
Pseudomonas syringae actinidiae Shaanxi M228 ANJI00000000
Pseudomonas syringae actinidiae Shaanxi M7 ANJJ00000000
Pseudomonas syringae aceris M302273PT AEAO00000000
Pseudomonas syringae aptata DSM 50252 AEAN00000000
Pseudomonas syringae atrofaciens DSM 50255 AWUI00000000
Pseudomonas syringae japonica M301072PT AEAH00000000
Pseudomonas syringae panici LMG 2367 ALAC00000000
Pseudomonas syringae pisi 1704B AEAI00000000
Pseudomonas syringae pisi HSE3 AUZR00000000
Pseudomonas syringae avellanae ISPaVe013 AKCJ00000000
Pseudomonas syringae avellanae ISPaVe037 AKCK00000000
Pseudomonas syringae 1212 AVCR00000000
Pseudomonas syringae 642 ADGB00000000
Pseudomonas syringae B64 ANZF00000000
Pseudomonas syringae pv B728a
Pseudomonas syringae FF5 ACXZ00000000
Pseudomonas syringae SM APWT00000000
Pseudomonas syringae theae ICMP 3923 AOJV00000000
Pseudomonas syringae theae NCPPB 2598 AGNN00000000
Pseudomonas syringae theae NCPPB 2598 ATDJ00000000
Pseudomonas syringae aesculi 0893 23 AEAD00000000
Pseudomonas syringae aesculi 2250 ACXT00000000
Pseudomonas syringae aesculi NCPPB3681 ACXS00000000
Pseudomonas syringae lachrymans M301315 AEAF00000000
Pseudomonas syringae lachrymans M302278PT AEAM00000000
Pseudomonas syringae mori 301020 AEAG00000000
Pseudomonas syringae morsprunorum M302280PT AEAE00000000
Pseudomonas syringae tabaci 6605 AJXI00000000
Pseudomonas syringae tabaci ATCC 11528 ACHU00000000
Pseudomonas syringae tabaci ATCC 11528 AEAP00000000
Pseudomonas syringae savastanoi NCPPB 3335 ADMI00000000
Pseudomonas syringae glycinea B076 AEGG00000000
Pseudomonas syringae glycinea R4 ADWY00000000
Pseudomonas syringae glycinea race 4 AEGH00000000
Pseudomonas syringae phaseolicola 1448A
Pseudomonas syringae phaseolicola 1644R AGAS00000000
Pseudomonas syringae maculicola ES4326 AEAK00000000
Pseudomonas syringae tomato K40 ADFY00000000
Pseudomonas syringae tomato Max13 ADFZ00000000
Pseudomonas syringae tomato NCPPB 1108 ADGA00000000
Pseudomonas syringae tomato T1 ABSM00000000
Pseudomonas syringae tomato DC3000
Pseudomonas viridiflava CC1582 AVDW00000000
Pseudomonas viridiflava TA043 AVDV00000000
Pseudomonas_viridiflava_UASWS0038_AMQP00000000

Figure 2. 9 Orthologs and *hrp* promoter motifs for DC3000 HrpL regulon orthologs in the *P. syringae* subgroup

A blank (white) cell indicates that no ortholog was detected. "==" indicates that an orthologous gene was identified but no upstream sequence could be extracted (due to incomplete sequence information and segmentation in draft genomes). For cases in which orthologs were detected and upstream sequences recovered, the color represents the – logarithm (base 10) of the HMM E-value for the best motif matching the *hrp* promoter model in the upstream sequence. A continuous color scheme is used where blue represents a poor match (E-value = 1), dark red indicates an intermediate match (E-value = 1e-02), and bright red indicates a good match (E-value 1e-05). Most verified *hrp* promoters in DC3000 match with values above 3. The leftmost gene column represents orthologs for the HrpL sigma factor, PSPTO_1404. In DC3000, this sigma factor is transcribed from a RpoN-responsive promoter [75]. Genes immediately downstream of *hrp* promoters are shown in columns, as they appear in *CEL*, *hrp/hrc* cluster, followed by type III effectors, chaperones and helpers, and non-type III function genes. Newly found members are in green background.

0……2……..5

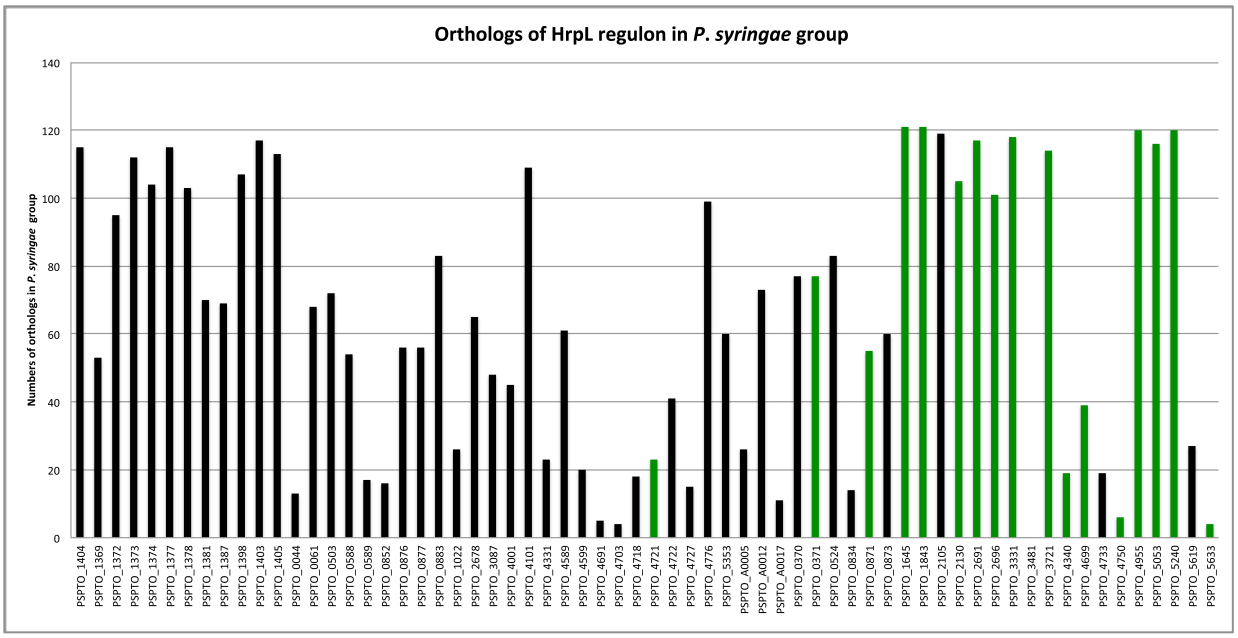


Figure 2. 10 Ortholog inventory of HrpL regulon in *P. syringae* group.

Green represents newly found members; black represents previously annotated regulon members. The values shown represent counts of orthologs of HrpL regulon members across 121 species.

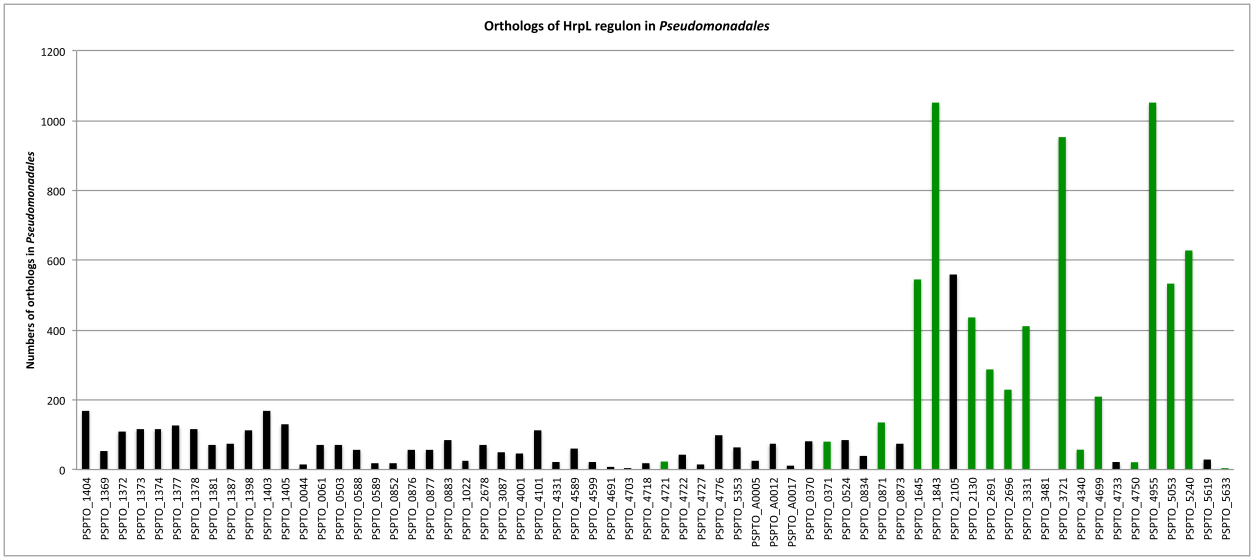


Figure 2. 11 Ortholog inventory of HrpL regulon in *Pseudomonadales*.

Green represents newly found members; black represents previously annotated regulon members. The values shown represent counts of orthologs of HrpL regulon members across 1060 species.

Discussion

Genome-wide approaches are well suited for investigating sigma factor regulons containing multiple chromosomally dispersed genes and operons [60, 115, 142, 149]. Several attempts have been made to define the set of genes regulated by HrpL in DC3000 [65, 80, 85] and these have succeeded in identifying many effectors and essential components of the T3SS. Although these efforts were intended to be thorough [86, 87], the availability of more advanced methodologies makes it both feasible and worthwhile to conduct a new inventory. The power of this approach was recently demonstrated in a report describing the use of RNA-Seq to link the expression of HrpL to members of the HrpL regulon in six different *P. syringae* isolates [130].

Using a combination of high-throughput sequencing and computational analyses, we have searched intensively for members of the HrpL regulon in DC3000 by identifying and confirming likely *hrp* promoters. The results increase the number of HrpL-responsive promoters in DC3000 to 73, including 52 of the 54 promoters already annotated in the DC3000 genome. The experiments reported here establish two important molecular details relevant to regulation by HrpL. First, candidate HrpL binding sites were defined by their immuno-co-precipitation with a FLAG-tagged version of HrpL, using both ChIP-Seq and more localized immunoprecipitation experiments to confirm the ChIP-Seq results. All of these regions contain a conserved motif that closely resembles the known *hrp* promoter sequence. Evidence for HrpL binding at these sites has not been reported previously. Second, we show that many of these motifs are active promoters. They are closely associated with transcription start sites, defined using RNA-Seq to capture mRNA 5'-ends that map to genomic coordinates immediately downstream of the putative promoters.

Using plasmid-based reporter systems, we demonstrate that cloned regions containing most of the promoter candidates also support transcription in a HrpL-dependent manner, and we can also detect HrpL-dependent transcription from regions downstream from the promoters in their native locations using qRT-PCR. Together, these experiments strongly support the addition of the new *hrp* promoters to the currently annotated set as well as the addition of the genes downstream of them to the HrpL regulon.

Integration of ChIP-Seq and RNA-Seq data with earlier DC3000 HrpL regulon inventories.

Many previously reported *hrp* promoters have been experimentally validated [65, 80, 85, 130]. Transcription start sites have also been globally mapped onto the DC3000 genome, including those associated with likely HrpL-dependent promoters [62]. Combining our data with these, we can summarize our current understanding of *hrp* promoters in DC3000 along with relevant evidence (Table 2.8 and Table 2.7).

All previously reported *hrp* promoters were confirmed with three exceptions. We saw no evidence for HrpL-binding upstream of PSPTO_3489 (sugar ABC transporter/ATP-binding protein), classified by Ferreira *et al.* [65] as a HrpL regulon member. A *hrp* promoter was also predicted within PSPTO_1369 (*shcN*) upstream of PSPTO_1370 (virulence factor HopN1) [65, 80, 85]. A TSS was detected three nucleotides downstream from this promoter (Figure 2.7), but other evidence suggests that PSPTO_1369 and PSPTO_1370 are co-transcribed from a *hrp* promoter upstream of PSPTO_1369, where HrpL binding is very clearly observed (Figure 2.7). Finally, although a *hrp* promoter was reported upstream of *PSPTO_3576* (*tvrR*, TetR-like virulence regulator) in an earlier 5'-capture experiment [62], we did not observe enrichment of the region containing it in our

ChIP-Seq screen. The activity of this promoter was also previously shown to be independent of HrpL in a promoter trap experiment [169].

We found evidence for three *hrp* promoters that had been noted previously but described as weakly supported. A putative promoter upstream of PSPTO_1645, reported by Ferreira *et al.* [65], was not included in their high-confidence list because the distance from the translational start site of the gene to the predicted promoter was too large. In contrast, we found evidence for HrpL-binding at this site and a TSS in close vicinity 3 to 7 bps from the 3'-end of the -10 region) (Table 2.7). The same authors report that a second gene, PSPTO_2691, was induced by HrpL but had no obvious *hrp* promoter associated with it. However, we detected HrpL binding, a captured TSS, and a *hrp* promoter motif upstream from this gene. PSPTO_2691 was also recently identified as a HrpL regulon member by Mucyn *et al.* [130]. Finally, the *hrp* promoter upstream of the *hopAT1* pseudogene (PSPTO_5618) was noted in [65] but was not included in the DC3000 annotation. Although the binding activity of HrpL at this location was not strong (Figure 4A), this region generated a positive result in the promoter fusion assay (Figure 2.5B) and transcription of the pseudogene was also elevated in the presence of HrpL in the qRT-PCR assay (Figure 2.5C).

A comparison between the results presented here and those reported recently by Mucyn *et al.* [130] reveals many similarities. Both studies identified the majority of known HrpL regulon members. In addition, we confirmed six previously classified HrpL regulated genes that were not detected by their survey (PSPTO_0044, PSPTO_1022, PSPTO_4101, PSPTO_4724, PSPTO_5353 and PSPTO_5616; Table 2.7). With respect to new regulon members, five are shared (PSPTO_0371, PSPTO_0871, PSPTO_2130, PSPTO_2691 and

PSPTO_4721). However, the two studies differ significantly when other new regulon candidates are examined. Fifteen new regulon members reported here (including PSPTO_5633; see below), and 12 in Mucyn *et al.*, are not in common. A likely explanation for the disagreement is that Mucyn *et al*. optimized their experiments to identify genes that respond to HrpL either directly or indirectly, and employed a cloned *hrpL* gene under the control of an arabinose-inducible promoter on a multi-copy plasmid. Some genes identified in this fashion would be expected to have no associated HrpL-binding activity or *hrp* promoter motifs in their upstream regions. This appears to be true for the 12 candidates reported by Mucyn *et al.* that are not in agreement (see note in Table 2.8). In contrast, the experiments reported here were optimized to identify genes directly regulated by HrpL. The *hrpL* gene was expressed from its native location, and candidates were identified by a combination of sigma factor binding, associated *hrpL* promoter motifs, TSS and the demonstration of HrpL-dependent transcriptional activity.

AprI (PSPTO_3331), an inhibitor of metalloprotease AprA, is regulated by HrpL

In *P. aeruginosa*, the protease AprA (orthologous to PSPTO_3332) is secreted by the T1SS and degrades flagellin monomers [170]. Recently, Pel *et al*. [171] analyzed this factor and its role in DC3000. Their results demonstrate that AprA is required for full virulence in DC3000 and functions by degrading flagellin before it can trigger plant defense mechanisms via the plant receptor FLS2. Although their work establishes that AprI can inhibit the protease either in vitro or when it is expressed in transgenic plants, they note that AprI is probably delivered to the bacterial periplasmic space by its T2SS signal peptide [172]. It is therefore unlikely to encounter the protease, making its function unclear.

Our results draw additional attention to this important system by establishing that *aprI* is downstream from a functional *hrp* promoter and is therefore a member of the HrpL regulon. The identity of the promoter is strongly supported by several lines of evidence (Table 2.8). An earlier transcriptome analysis identified a captured 5'-end and a potential RpoN promoter motif at this position [62] but we are unable to detect RpoN binding to this region in a ChIP experiment (data not shown). Filiatrault *et al*. also observed a captured 5'-end upstream from *aprA*, possibly linked to an RpoD promoter. The two genes, which are 59 bps apart, therefore appear to be regulated independently. Why *aprI* is regulated by HrpL is not obvious, especially if it is expressed at the same time as *aprA* (i.e., during an infection). It is possible that the target for AprI in the periplasm is another protease unrelated to AprA. However, deletions in *aprI* in either DC3000 or ΔQ1-1 backgrounds exhibit no growth or virulence phenotypes *in planta*, leaving the role of AprI a mystery. AprI orthologs are widespread in the *P. syringe* group (Figure 2.9) although in some cases without accompanying *hrp* promoter motifs. It is also found in other pseudomonads including *P. fluorescens* (Table 2.7).

PSPTO_5633 (HopBM1) is a weak effector

Three characteristics common to most known *P. syringae* effectors are transcription via HrpL, the appearance of effector homologs in other pathogens, and an N-terminal protein sequence with certain characteristic features [167]. The N-terminal features include multiple serine residues within the first 50 amino acids, an aliphatic residue (isoleucine, valine, leucine, alanine, methionine) or proline at positions 3 or 4, and a lack of acidic amino acids (aspartic acid, glutamic acid) within the first 12 residues [167]. While

new regulon members such as PSPTO_3331 (also discussed above), PSPTO_4340 (insecticidal toxin protein) and PSPTO_4699 (a non-ribosomal peptide synthetase component) satisfy the N-terminal criteria for T3SS substrates, their annotated functions suggest that they are not effectors.

PSPTO_5633 did not appear in the original DC3000 annotation but was added following a global transcriptome analysis [131]. PSPTO_5633 shares high sequence similarity with hypothetical proteins in three members of the *P. syringae* group (Figure 2.9) as well as with proteins in several other pathogenic bacterial species (mentioned above). However, its leader sequence scores poorly when it is examined for the amino acid patterns mentioned above [167]. In addition, DC3000 plasmid gene PSPTO_B003, identical in sequence to PSPTO_5633, was reported to have no T3SS translocation activity in a *P. fluorescens* background [173]. Our data demonstrate that PSPTO_5633 is detectably but weakly translocated into plant cells in a manner that depends on the T3SS (Figure 2.8B). The disagreement between these results is probably due to the fact that translocation in general is lower in the *P. fluorescens* system than in DC3000 [173]. Since PSPTO_5633 translocates poorly in DC3000, its translocation may be undetectable in *P. fluorescens*.

An analysis of the N-terminal region of PSPTO_5633 using SignalP [174] suggests that it may be secreted through the Sec pathway (Figure 2.8C). Other proteins within the HrpL regulon share this characteristic, including PSPTO_3331 (newly found), PSPTO_0524, and hopAJ1. Our results are not consistent with a model in which PSPTO_5633 is delivered through the T2SS, either directly or indirectly, into plant cells (Figure 2.8B). However, since this experiment depends on the interaction of Cya with calmodulin within the plant cell, it

does not test secretion through the T2SS *per se*. Additional experiments will be required to determine whether PSPTO_5633 or other HrpL regulon members are T2SS substrates.

HrpL regulon members are represented in other Pseudomonadales

The new HrpL regulon members are largely conserved within the *P. syringae* group, albeit at varying levels (Figure 2.9). Table 2.4 summarizes the orthology analysis for all 1060 *Pseudomonadales* genomes. Although the large size of this data set precludes a detailed analysis, broad patterns can be discerned. Orthologs for the core regulon components such as *hrpL*, *CEL*, *hrp/hrc* and effector genes are infrequently found outside the *P*. *syringae* group. However, several new regulon members are widely represented (Figure 2.11). Examples are PSPTO_1843 (aspartate kinase), PSPTO_3721 (enoyl-[acyl-carrier-protein] reductase) and PSPTO_4955 (thiosulfate sulfurtransferase / phosphatidylserine decarboxylase) whose orthologs appear in almost all sequenced strains tested. Interestingly, in some cases these are accompanied by upstream *hrp* promoter motifs even in genomic contexts where HrpL is absent, although this is relatively uncommon.

Diversity in the number of HrpL regulon orthologs and their distinct patterns of conservation across the *Pseudomonadales* imply that HrpL regulon member recruitment is a complex process. A gene is likely to be selected for HrpL regulon membership if its expression confers an advantage in the nutrient-poor and stressful environment of the plant apoplasm, especially when coupled to the expression of the regulon as a whole. The assembly and refinement of a working HrpL regulon probably arises due to multiple evolutionary events, including horizontal gene transfer [120], changes in coding regions

that alter or eliminate protein function (e.g. conversion into pseudogenes), and variations in *cis*- acting elements [130] or the factors that recognize them. Promoter alteration may be the easiest recruitment mechanism for inclusion into a regulon [175].

In summary, our analysis revealed 20 new HrpL regulon members. The combination of laboratory and computational methods used here makes it likely that the inventory at this point is nearly complete. One conceptual difficulty in "closing" the list is that relaxed criteria can result in candidates that satisfy one of several tests for membership (such as the apparent HrpL-dependent transcription from a region directly upstream from PSPTO_0816 in a promoter trap assay). We suggest however that evidence for HrpL binding, the presence of a motif, and demonstrated transcriptional activity together best define the regulon in practical terms. The remaining challenge is to determine what roles, if any, the new regulon members play in the process of pathogenesis.

Acknowledgements

We would like to thank Magdalen Lindeberg for discussion concerning novel HrpL regulon members and their orthologs. Christopher R. Myers, Alan Collmer, John Helmann, Keith Perry, and Melanie J. Filiatrault provided helpful feedback on the manuscript. Melanie J. Filiatrault also assisted in the design of the RNA-Seq and 5' capture protocol. We thank Brian H. Kvitko for providing pCPP5388.

CHAPTER 3

CHARACTERIZATION OF SIGMA FACTOR RPON REGULON IN *PSEUDOMONAS SYRINGAE* PV. *TOMATO* DC3000

Abstract

RpoN (σ^{54}) is the central regulator for many bacterial cellular processes. Because enhancer-binding proteins are required to activate RpoN dependent transcription, the RpoN regulon has the capacity to fine-tune its transcriptional output in response to environmental signals. In *Pseudomonas syringae* pv. *tomato* DC3000, it is known that the T3SS requires RpoN for expression. However, little is known about the RpoN regulon otherwise. Using ChIP-Seq and RNA-Seq, we identified candidate RpoN regulated genes, and also those that are differentially expressed in *hrp* inducing versus non-inducing conditions. We identified 226 regions that were enriched by the ChIP-Seq procedure, among which 214 contain at least one RpoN promoter-like motif. The concensus sequence closely resembles the known RpoN-responsive promoter element, 5'-TGG-N9-TGC-3'. Based on these results, RpoN appears to bind to promoters upstream from a variety of genes involved in flagella biosynthesis, energy metabolism, nitrogen metabolism, and transport. RpoN also controls the transcription of small noncoding RNAs, enhancer binding proteins, and other regulatory proteins including (possibly) itself. 225 genes are differentially expressed in *hrp*-inducing conditions when compared with expression in a non-inducing medium. Some of the up-regulated genes, including amino acid ABC transporters (PSPTO_1255 and PSPTO_4171), are likely candidates for genes involved in

bacterial infection. PSPTO_0927 (type IV pilus protein), required for virulence in *P. syringae* pv. *tabaci* [176], and Hcp-1 and Hcp-2 (secreted proteins which are possibly components of the Type VI secretion system, known to be involved in host infection of *P. aeruginosa* [110]) are also potentially regulated by RpoN. Our data suggests that RpoN does not directly regulate the genes responsible for coronatine biosynthesis but might regulate up to three genes in the T3SS including *hrpL*. Our results significantly improve our understanding of the pathways in DC3000 that are controlled by RpoN directly or indirectly.

Introduction

RpoN is best known for its regulation of nitrogen fixation and assimilation [91, 100], carbon source selection, and energy metabolism [113, 114, 177, 178]. In many bacteria, flagella biosynthesis requires RpoN to make components of the motility machinery [113, 115, 177-185]. Swarming [186], and quorum sensing [177, 187, 188] are also well known to be regulated by RpoN. Besides T3SS, RpoN also regulates the T4SS [104, 110, 113] and T6SS [110, 115, 189]. Resistance to oxidative stress, osmotic stress, acid, and starvation is also controlled by RpoN [178, 184, 190, 191]. Biofilm formation is dependent on RpoN in *Vibrio anguillarum* M3 and *Burkholderia cenocepacia* [112, 192]. Importantly, RpoN also regulates sigma 70 family proteins such as RpoS and RpoF (FliA) [180, 181, 193], indirectly regulates AlgU by regulating the sensor kinase KinB [177], and regulates small RNA *flaX* involved in motility [115].

RpoN controls virulence in many bacteria. For example, it regulates *hrpL* transcription, and consequently controls virulence, growth *in planta*, the hypersensitive

response, and the expression of *hrp* genes in *Erwinia amylovora* [111, 194]. RpoN is also required for virulence of fish pathogens such as *Vibrio anguillarum* [112], and *Edwardsiella tarda* [193]. Coronatine production and HrpL expression were suggested to be controlled by RpoN in *P. syringae* pv. *glycinea* [195].

RpoN-mediated transcription in bacteria shares some features with transcription in eukaryotes. The requirement by RpoN for enhancer binding, ATP for melting the closed complex, and chromosome bending, is reminiscent of eukaryotic polymerase II-mediated transcription. In contrast, bacterial transcription mediated by sigma 70 subunits proceeds without these factors [196]. RpoN is also very distinct structurally from sigma 70 family sigma factors [197]. These differences place RpoN into a unique class (σ^{54}), easily distinguished from the σ^{70} family. RpoN usually requires two factors to activate transcription, namely, integration host factor (IHF) [103, 104], and EBPs [105, 106]. IHF facilitates DNA bending to loop EBPs, which are usually bound at a distance from the RpoN binding site, back to the σ^{54}-closed complex (RNA polymerase holoenzyme). The involvement of enhancer binding proteins allows RpoN to independently regulate multiple targets, presumably in response to various cues.

The mode of transcription initiation by RpoN and its EBPs is also unlike that for other bacterial sigma factors. RpoN, along with the rest of the RNA polymerase, can bind to DNA stably without transcription [198]. Transcription activation requires enhancer binding to an UP element region surrounding the RpoN binding site and also requires adenosine triphosphate (ATP).

How EBPs function to activate bacterial transcription has been summarized by Bush and Dixon [107]. EBPs usually have a regulatory domain to repress oligomerization until

activated or to promote the formation of multimers in response to signals. Some EBPs are in preformed hexamers presumably enabling them to respond quickly to environmental changes. Upon forming hexamers, EBPs can weakly associate with σ^{54}-holoenzyme. After binding ATP, the central EBP domain (the AAA+ domain involved in ATP binding and hydrolysis) becomes active, establishes a strong connection with σ^{54}-holoenzyme, and changes the conformation σ^{54} and other components of the holoenzyme. These changes facilitate transition from a closed transcription initiation complex to an open complex [108].

The role of RpoN has been studied in only a few pseudomonads. In *P. aeruginosa*, RpoN controls pilin formation [199], quorum sensing [188], the synthesis of flagellin, extracellular polypeptides, outer membrane proteins and the alginate capsule [200]. It is also involved in tolerance to antimicrobials [201]. RpoN also regulates a sensor kinase, KinB, which regulates virulence factors important for disease development in *P. aeruginosa* [117]. The role of RpoN in *P. aeruginosa* PA14 appears to be host-dependent. RpoN mutants are less virulent in mouse, and have reduced ability to kill *Caenorhabditis elegans*, but still cause disease symptoms similar to wild-type in *Arabidopsis*, and are not impaired in virulence in the greater wax moth insect *Galleria mellonella* [202].

In *P. putida*, RpoN mutants are nonmotile and cannot utilize many nitrogen sources such as nitrate urea or uncharged amino acids [203]. In *P. syringae* pv. *maculicola*, RpoN affects multiplication and disease formation in the host by regulating transcription of *hrpL* [68, 204]. In *P. fluorescens* CHAO, RpoN controls antibiotic production, utilization of carbon and nitrogen sources and the ability to protect cucumber from a root rot pathogen *Pythium ultimum* [205]. In *P. fluorescens* SBW25, RpoN is required for colonization of sugar beet

seedlings [178]. RpoN is also required for root colonization in *P. chlororaphis* [206]. In *P. syringae*, RpoN was shown to control coronatine [195].

In many bacteria, it is well established that flagella biosynthesis is regulated by RpoN [113, 115, 177-183, 185]. Flagella are extremely complex structures consisting of a rotary motor embedded in the membrane and an external flagella filament extending from the cell (Figure 3.1); see also references [207, 208] for recent reviews of this system in bacteria. Flagella are thought to be important in DC3000 by providing the means by which the bacterial cells propel themselves through stomata into the leaf apoplast [27]. Flagellin protein, a component of the flagellar filament, is also important in pathogenicity of a fish pathogen *Edwardsiella tarda* [209].

Flagella share similarity in structure with components of the T3SS [210, 211], and in fact flagellin can be transported through the T3SS [36]. Some anti-virulence compounds, such as salicylidene acylhydrazides, target both flagella and T3SS but the mechanism is unknown [212]. The flavonoids, one of the most abundant plant secondary metabolites in plants, were shown to target GacA/S and affect both motility and expression of *hrpL* [213].

In *P. syringae* pv. *tomato* DC3000, the ability to colonize plants and subdue multiple layers of plant defense is dependent on the T3SS [119]. HrpL is a direct regulator of the T3SS [71, 128] and T3SS expression is predicted to be tightly controlled by HrpL [214]. In turn, *hrpL* transcription is dependent on RpoN and the HrpR/S EBPs [49]. HrpR/S, a heterodimer enhancer required for activation of *hrpL*, has an RpoN interacting domain and a DNA binding domain but lacks a response regulatory domain to transduce environmental signals [75, 215]. The environmental cues may influence HrpR/S via other intermediary factors. DC3000 encodes more than a dozen enhancer protein candidates [107] but the

regulatory function (in terms of promoter activation) is unknown for nearly all of them. Figure 3.2 summarizes known and potential pathways regulated by RpoN.

The RpoN motif is highly conserved across bacterial species [103]. In contrast to the typical RpoD-type promoter with its –35 and –10 regions, RpoN recognizes two regions spaced at -24 (TGG) and -12 (TGC) relative to the transcription start site. Although motif conservation makes it possible to predict RpoN promoters in DC3000 using a profile HMM, this approach is "noisy" because it does not account for other contextual clues in the genome such as the presence or absence of upstream enhancer binding sites, accompanying transcription start sites, and the distance to nearby genes. Without verified RpoN binding and promoter activity, it is difficult to distinguish between genuine and false positives.

In this work, we focus on three questions. (i) Does RpoN regulate genes within the *hrp* regulon, or does it act primarily through HrpL? (ii) Can we define the DC3000 RpoN regulon and determine if RpoN regulates other factors potentially involved in virulence and pathogenicity besides *hrpL*? (iii) Can we detect RpoN activation of specific genes in response to a shift in culture medium (*hrp*-repressing to *hrp*-inducing conditions)? To address these questions, we used ChIP-Seq and RNA-Seq (coupled with a 5' capture procedure) to identify candidate RpoN regulated genes and differentially expressed genes in *hrp* inducing conditions.

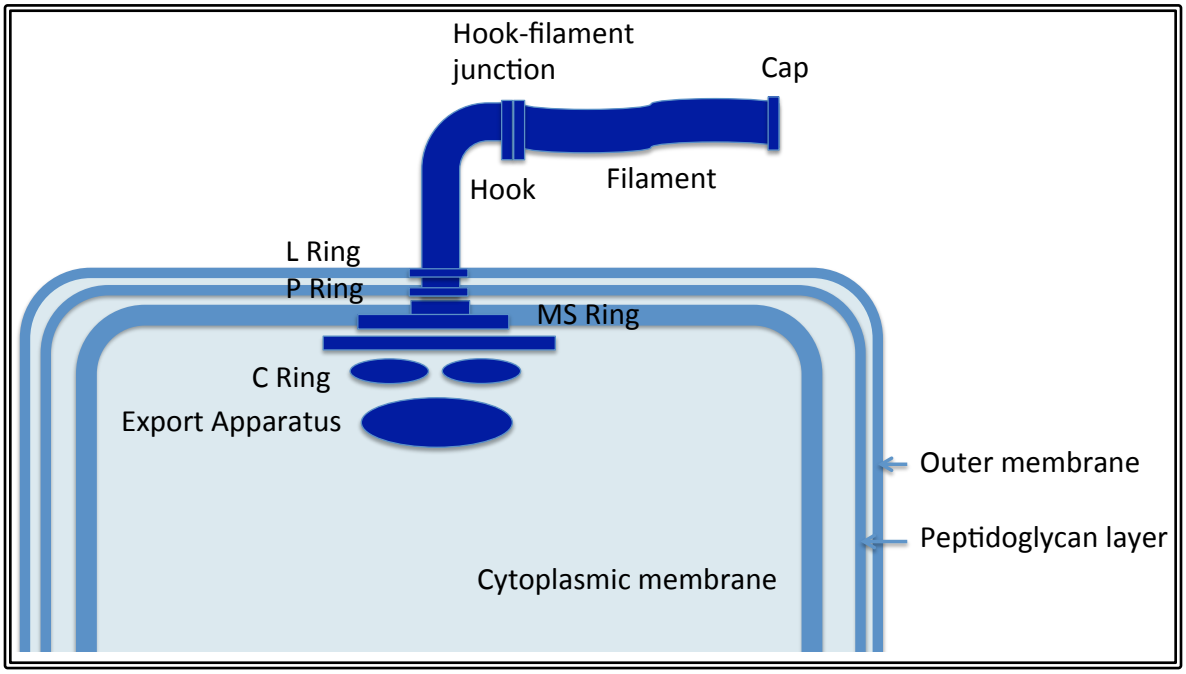


Figure 3. 1 Schematic diagram of flagella structure in gram-negative bacteria

A flagellum consists of three parts: 1) the basal body containing export apparatus, MS and C ring (rotor), the rod and LP ring, 2) the hook and 3) the filament. Figure is simplified from reference [216].

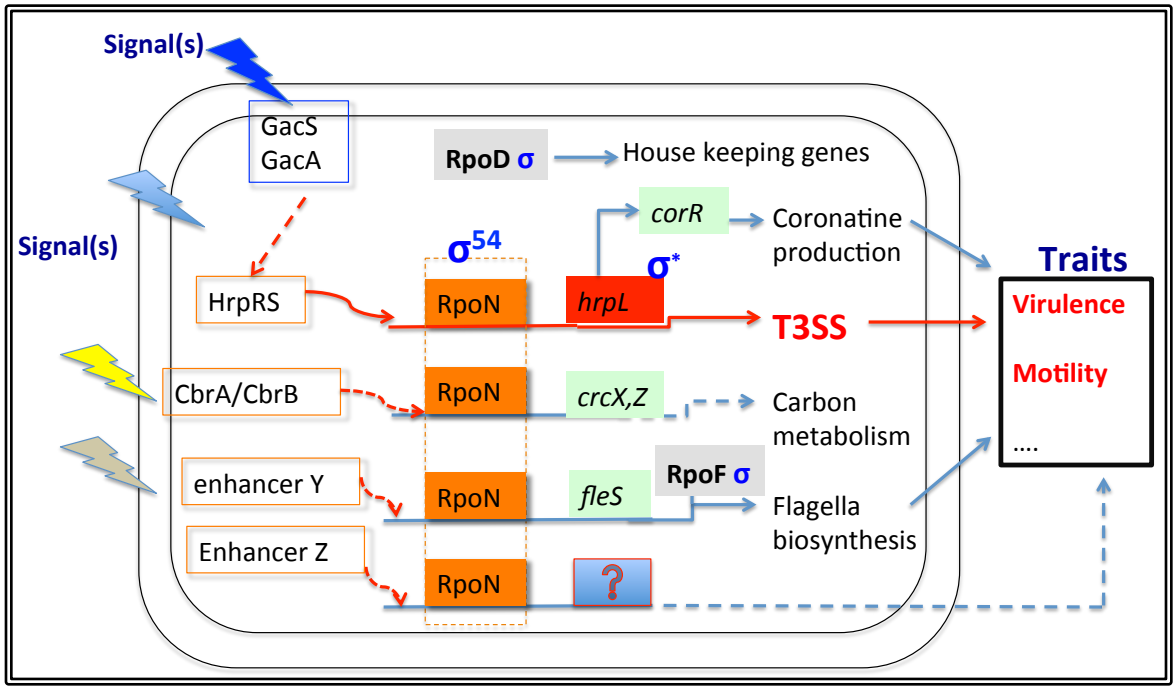


Figure 3. 2 RpoN as a master regulator in bacteria.
RpoN (σ^{54}) regulates T3SS (and coronatine production via HrpL), carbon metabolism, flagella biosynthesis and many other processes, all of which contribute to macroscopic traits including virulence, and motility. RpoN requires EBPs for transcriptional activation, which allows bacteria to integrate environmental cues to modulate physiological growth and response.

Materials and methods

Bacterial strains and growth conditions

DC3000 and derivatives were grown at 28°C in Kings B (KB) medium [135] or KB medium solidified with 1.5% (wt/vol) agar. *Escherichia coli* was used as the host for plasmid manipulations. *E. coli* was grown at 37°C in Luria-Bertani (LB) [137] medium or LB medium solidified with 1.5% (wt/vol) agar. Rifampicin, and kanamycin were used at 50 μg/ml, and 50 μg/ml, respectively. Bacterial strains used in this experiment, along with their relevant antibiotic resistance markers, are listed table 3.1.

Bacterial growth and sample collection for RNA-Seq and ChIP-Seq were carried out as follows. Colonies of *rpoN-FLAG* (see below for construction) were obtained from KB plates that had been incubated for 48 hours. Cells were re-suspended and grown overnight in 6 x 150 ml KB at 28°C in a shaking incubator at 250 rpm. The overnight culture was then pelleted by centrifugation and washed in KB and MG media [134] for samples to be transferred to KB and MG media, respectively. For the non-inducing condition, washed cells were re-suspended in KB and inoculated into bioreactors (Infors-HT, Switzerland) containing 400 ml KB. For the inducing condition, washed cells were re-suspended in MG and inoculated into bioreactors containing 400 ml MG medium supplemented with iron citrate (Sigma, St Louis, MO) to a final concentration of 50 μM as described previously [134]. Five milliliters and 100 ml of samples were collected for RNA-Seq and ChIP-Seq, respectively, at 1.5 hours after the medium change. Culture samples for RNA-Seq were combined with two volumes of RNAprotect Bacteria (Qiagen) to stabilize RNAs and were stored at -70^0C before RNA extraction. Samples for ChIP-Seq were cross-linked with 37% Formaldehyde (1% final concentration) for 20 minutes. Cross-linking was quenched with

2.5 M Glycine (0.36 M final concentration). Cells were collected and washed twice with Tris-buffered saline (TBS). Washed pellets were stored at -70^0C until processing. Three biological replicates were collected for each condition for the RNA-Seq experiments. One sample was collected for the ChIP-Seq experiment for each condition.

Transformations and genomic modifications

Suicide vectors for single-crossover insertions were introduced into DC3000 backgrounds using electroporation [145]. Gene replacements using pK18mobsacB were performed as described previously [146]. Plasmid insertions using pK18mobsacB were selected by plating on KB medium with kanamycin. Plasmid integration was confirmed by PCR, antibiotic resistance and sequencing.

Construction of rpoN-FLAG. Regions flanking the *PSPTO_4453 (rpoN)* gene were amplified by primers HL273/HL274 and HL275/HL276 (see Table 3.2 for all primer sequences) from DC3000 genomic DNA, purified by gel electrophoresis (Qiagen), and joined by SOEing PCR [147]. The joined fragment was then digested with XbaI, and ligated with XbaI digested pK18mobsacB. The resulting pK18mobsacB construct was sequenced using primers M13F, M13R, HL277, HL278, HL279, HL280, and HL281 to ensure that no unexpected mutations had been introduced in the manipulated region. The FLAG-tagged construct was transformed into DC3000 by electroporation. The merodiploid intermediates were selected for growth on medium containing kanamycin. Recombinants that had eliminated pK18mobsacB plasmid sequences were identified by sucrose counter-selection. Sucrose-resistant, Km-sensitive colonies were confirmed by Sanger sequencing with primers

HL277, HL278, HL279, HL280, HL281, HL282 and HL283. The resulting *rpoN FLAG*-tagged *P. syringae* pv. *tomato* DC3000 is referred to as *rpoN-FLAG* hereafter.

Chromatin immunoprecipitation with exonuclease treatment (ChIP-exo) paired with high throughput sequencing.

The bacterial strain *rpoN-FLAG* was used to map RpoN binding sites. Binding activity was studied in two conditions, growth in KB or MG supplemented with ferric citrate. ChIP-exo libraries for high throughput sequencing were prepared as described in chapter 2.

RNA Isolation and preparation for RNA-Seq.

Total RNA was prepared using a miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions, using the optional on-column DNaseI digestion. RNA was treated twice with DNase I (Ambion) to remove residual DNA and then cleaned and concentrated using RNA cleanup and concentrator-5 (zymoresearch). Integrity of the RNA was assessed using Bioanalyzer (Agilent technologies, Cornell University Life Sciences Core Laboratory Center Microarray Facility, Cornell University). Only RNAs with high quality were used for further analyses.

Depletion of processed RNAs, ligation of tag sequence and construction of cDNA libraries for 5' mapping and RNA-Seq.

RNA samples were prepared for RNA-Seq as described in chapter 2.

Table 3. 1 Bacterial strains and plasmids used in this study

Designation	Genotype and Relevant Features	Reference
Plasmids		
pK18mobsacB	Small mobilizable suicide vector, sucrose-sensitive (sacB)/ Kan^R	[138]
pHL1	Destination vector for transcriptional fusion using LR reaction, carrying *gfp*/ Tet^R Cam^R	This study (HrpL regulon)
pBS181	p*hrpJ*::iucD expresses GUS under T3SS inducing conditions/ Tet^R Kan^R	This study (HrpL regulon)
E.coli		
TOP10	Φ80*lacZΔM15 ΔlacX74*	Invitrogen
Pseudomonas syringae pv. tomato		
DC3000	*Pseudomonas syringae* pv. *tomoto* DC3000/ Rif^R	[129]
HLN394	*Pseudomonas syringae* pv. *tomoto* DC3000 with RpoN C-Terminal FLAG-tagged at its native genomic locus/ Rif^R	This study
HLN395	*Pseudomonas syringae* pv. *tomoto* DC3000 with *rpoN* deleted (Δ*rpoN*)/Rif^R, $Spec^R$	[217]
HLN201	Δ*pvsA* carrying *hrpL* coding region without stop codon fused with *gfp*/ Rif^R Tet^R Kan^R	This study
HLN202	Δ*pvsA* carrying region from -23 base pairs upstream of *hrpL* to the end of HrpL coding region without stop codon fused with *gfp*/ Rif^R Tet^R Kan^R	This study
HLN203	Δ*pvsA* carrying region from -111 base pairs upstream of HrpL to the end of *hrpL* coding region without stop codon fused with *gfp*/ Rif^R Tet^R Kan^R	This study
HLN204	Δ*pvsA* carrying region from -166 base pairs upstream of *hrpL* to the end of *hrpL* coding region without stop codon fused with *gfp*/ Rif^R Tet^R Kan^R	This study
HLN205	Δ*pvsA* carrying region from -315 base pairs upstream of *hrpL* to the end of *hrpL* coding region without stop codon fused with *gfp*/ Rif^R Tet^R Kan^R	This study
HLN398	*P. syringae* DC3000 carrying *hrpJ* promoter upstream of *iucD* (GUS reporter gene)/ Rif^R $Spec^R$ Tet^R Kan^R	This study
HLN399	*rpoN-FLAG* carrying *hrpJ* promoter upstream of *iucD* (GUS reporter gene)/ Rif^R $Spec^R$ Tet^R Kan^R	This study
HLN400	Δ*rpoN* carrying *hrpJ* promoter upstream of *iucD* (GUS reporter gene)/ Rif^R $Spec^R$ Tet^R Kan^R	This study

Table 3. 2 Primers used in this study

Names	Sequence	Function
Primers for making *rpoN-FLAG*		
HL273/ HL274	TATTATCTAGAACGCCATCACCATTCAG TATTACGC/ CTTGTCATCGTCGTCCTTGTAGTCCATG GCAGGGGCTAAACACCTTA	Forward/reverse primer for amplification of flank A of *rpoN*-FLAG construct
HL275/ HL276	ATGGACTACAAGGACGACGATGACAAG AAACCATCGCTAGTCTTGAGAAT/ ATTATCTAGAGCTCGGCCTGCATCACAT CAATCGC	Forward/reverse primer for amplification of flank B of *rpoN*-FLAG construct
HL277	CGATTCTCGAAACCCGCAAGGA	Forward primer for checking the whole manipulated region of RpoN surrounding flank A
HL278	GGCAGTATCGACCGGCAGTTCATT	Reverse primer for checking the whole manipulated region of RpoN surrounding flank A
HL279	GCTGGACATCTGCGAAACCG	Forward primer #2 for checking the whole manipulated region of RpoN surrounding flank A
HL280	CCTGACGCAGCTCGTCTTCCTTG	Reverse primer #2 for checking the whole manipulated region of RpoN surrounding flank A
HL281	CGTGCATGCCTACCGCTTCG	Reverse primer #3 for checking the whole manipulated region of RpoN surrounding flank A
HL282/ HL283	AGGGCGAAGTCGATGTGTTTACC/ TACAGTTTTCTCCTTTGCCAGTGC	Forward/reverse primer for checking the whole manipulated region of RpoN surrounding flank B
RT-PCR primers in coding region		
HL298/ HL299	GCAACAGGCAATCCGCTTACTTC/ CGTTCCAGCATAGGGTTTGACTCC	Forward/reverse RT-PCR primer for coding region of *rpoN*
HL300/ HL301	CGGAGCAGTCAATAAGATGG/ TACAGGTATCGGTAGTCAGG	Forward/reverse RT-PCR primer for coding region of PSPTO_0927
HL302/ HL303	CTTAACGACCTGACCGAAACCC/ CGCCAGAGAAGCAACCATCC	Forward/reverse RT-PCR primer for coding region of PSPTO_1955
HL304/ HL305	GCAGAACGGCACTGTTGACG/ AGCAGGCGTGTACCGACTTC	Forward/reverse RT-PCR primer for coding region of PSPTO_4171
HL306/ HL307	ACGACACCCATAACAAGAAC/ CCCAAATACAACTCCTCTCC	Forward/reverse RT-PCR primer for coding region of PSPTO_5668
HL308/ HL309	CTCTGTGTTTCCGACCATTCTGC/ GACCGTCCTGACCGTGAAGC	Forward/reverse RT-PCR primer for coding region of PSPTO_1976
RT-PCR primers at potential RpoN binding sites		
HL240/ HL241	TAAGCCCATGTTTCAGAAGATTGTG/ CCGTCATTTGGCGAATACCG	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_1404
HL242/ HL243	AAGGTGTTCTGACGGATACGC/ GCATTCATTCTTACTCTCCAAGGC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_0927
HL244/ HL245	CGGAGGATTGACGCCTGTTTC/ TCAGTAAACGGTGTGTTAAGAGAGAC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_1955

HL246/ HL247	GGCACTGGCTTTGTAAAAGTTGG/ GCGATCCACTGATTCGTTACCTG	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_1976
HL248/ HL249	ATGGCACTTTTCTGTCATCTTTGTAGC/ CGTTCAGCGTCGGGTGGATTC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_4936
HL250/ HL251	ACATCTATCCAGCAAGGAGCAC/ ACCGAGTCAGCAGTGAAGGC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_5435
HL290/ HL291	ATCATGGGCGACTCTCAGCAG/ CAACGGCAACTGGCACGATTC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_3331
HL292/ HL293	CATTCAGGAACCGACCGCTTC/ TTGATAGAGCCGCAATGAGAGG	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_4171
HL294/ HL295	GCTTCTGCTATATGCTTAG/ CCGTGTTGTTCTTGTTATG	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_5668
HL296/ HL297	CGTAACGGCTCTGACATAACAAG/ CCGAAAGGTCCCTGTAAAGC	Forward/reverse RT-PCR primer at ChIP-Seq peak upstream of PSPTO_5669

Alignment of Illumina reads to the DC3000 genome sequence and Profile generation

The alignment procedure is similar to the one described in chapter 2, except that the first 43 nucleotides of each read with quality score of 20 or above (99% of inferred base calls are accurate) were aligned to the reference genome of DC3000 (accession number: AE016853) using Bowtie [218]. Subsequent analyses focused on the main chromosome of DC3000.

Motif detection using MEME

The three replicate 5' capture profiles were combined to make a single profile for each condition. A promoter motif search was conducted using the top 1000 'peaks', which are the 1000 genome coordinates with the highest number of associated read counts in the profiles. Because transcriptional start sites are generally within a short distance from promoters, 40 bases upstream from each of the top peak coordinates were extracted. Then MEME [219] was used to identify over-represented motifs with the following parameters:

-dna –mod anr –nmotifs 20 –minw 14 –maxw 35

For identification of RpoN motifs in vicinity of enriched ChIP-Seq regions, DNA sequences ranging from 60 to 80 bases in length were extracted from the 'peak regions' such that the extracted sequences straddled the center of the peak region. The collection of DNA sequences were subjected to MEME with the same parameters as above, except both strands were examined:

-dna –mod anr –nmotifs 20 –minw 14 –maxw 35 – revcomp

Identification of regions enriched by ChIP-Seq.

Enriched sequence reads were identified using Genetrack as described previously [153] and in chapter 2. Since the Genetrack output for the two conditions (KB and MG supplemented with iron) was identical, the two profiles were combined before final Genetrack analysis, resulting a single list of RpoN-binding sites. Regions of 60 to 80 bps surrounding the RpoN-binding sites were searched for motifs using MEME as described above. Genes potentially regulated by RpoN were functionally classified using the comprehensive microbial resource database downloaded from the J. Craig Venter Institute (JCVI) [220].

Identification of Differentially Expressed Genes using DE-Seq.

Sinister profiles [131] were used to assign counts to all operons in DC3000. Read counts for all genes for 3 replicates from two conditions (KB and MG with iron) were analyzed using DE-Seq [221]. Low count genes were not included in the analysis (a 0.4 quantile cutoff was applied to counts for each gene summed across all conditions, and the genes with the lowest 40 % quantile were removed). Differentially expressed (DE) genes with p-value cutoff of 0.001 are reported here. DE genes were classified into role categories using the database downloaded from JCVI [220].

qRT-PCR and qPCR

cDNA synthesis was accomplished using qScript cDNA Synthesis (Quanta, Biosciences). qPCR steps were performed using iQ SYBR Green Supermix (Bio-Rad). Primer pairs amplifying a DNA fragment of length around 100 base pairs were designed by Beacon

Designer™. For confirming enrichment of ChIP-Seq binding sites, primers were designed to amplify 100 base pair regions surrounding the putative targets. To test for transcript abundance, coding regions were amplified.

Swarming assay

Swarming assays were conducted as described previously [222]. Briefly, plates for swarming assays were prepared using 8% (wt/vol) NB medium solidified with 0.5% (wt/vol) agar. A volume of 5ul of bacterial suspension (OD_{600} 0.3) was spotted onto the agar plate. Dry plates were incubated at 25^{o}C, and photographs were taken after 2 days.

Hypersensitive response assay

Solanum lycopersicum plants were germinated and grown in a greenhouse with approximate 16/8 hour light/dark cycles. Four to five week old plants were inoculated with a 3 x 10^7 CFU/ml bacterial suspension using blunt syringe infiltration. Hypersensitive response was observed and documented by photography after 2 days.

Results and Discussion

RpoN binding sites identified by ChIP-Seq

To perform a global inventory of genomic sites likely to bind RpoN in D3000, we first modified the *rpoN* locus at its native position in the DC3000 chromosome (PSPTO_4453) to fuse a FLAG epitope to the 3' end of the gene. The functionality of the tagged protein was tested in several ways before conducting a ChIP-Seq analysis. The *RpoN-FLAG* strain retained the ability to stimulate the hypersensitivity response when

inoculated into tomato plants (Figure 3.3A), which confirmed that the modified strain is able to deploy the T3SS. Using a promoter trap assay, activation of a specific HrpL-dependent promoter was also observed in the *rpoN-FLAG* background in response to growth in *hrp*-inducing medium (Figure 3.3B). Swarming, another response that is dependent on RpoN, was also observed in a plate assay (Figure 3.3C), and transcriptional activation of *hrpL* was detected using qRT-PCR when cells were shifted from KB to MG medium (Figure 3.3D). Finally, using chromatin immunoprecipitation followed by PCR, enrichment was detected at multiple candidate RpoN binding sites (Figure 3.3E). As expected, only the *rpoN-FLAG* sample showed enrichment. These preliminary experiments provide confidence that *rpoN-FLAG* behaves in a physiologically normal fashion.

A ChIP-Seq analysis was conducted using cells grown in KB (a *hrp*-repressing rich medium) and MG supplemented with iron (a *hrp*-inducing medium). Using the ChIP-exo procedure adapted from [148] and described in chapter 2 of this thesis, we prepared DNA samples for high-throughput sequencing. Table 3.3 summarizes the mapping statistics of the two libraries using data generated by the Illumina Genome Analyzer.

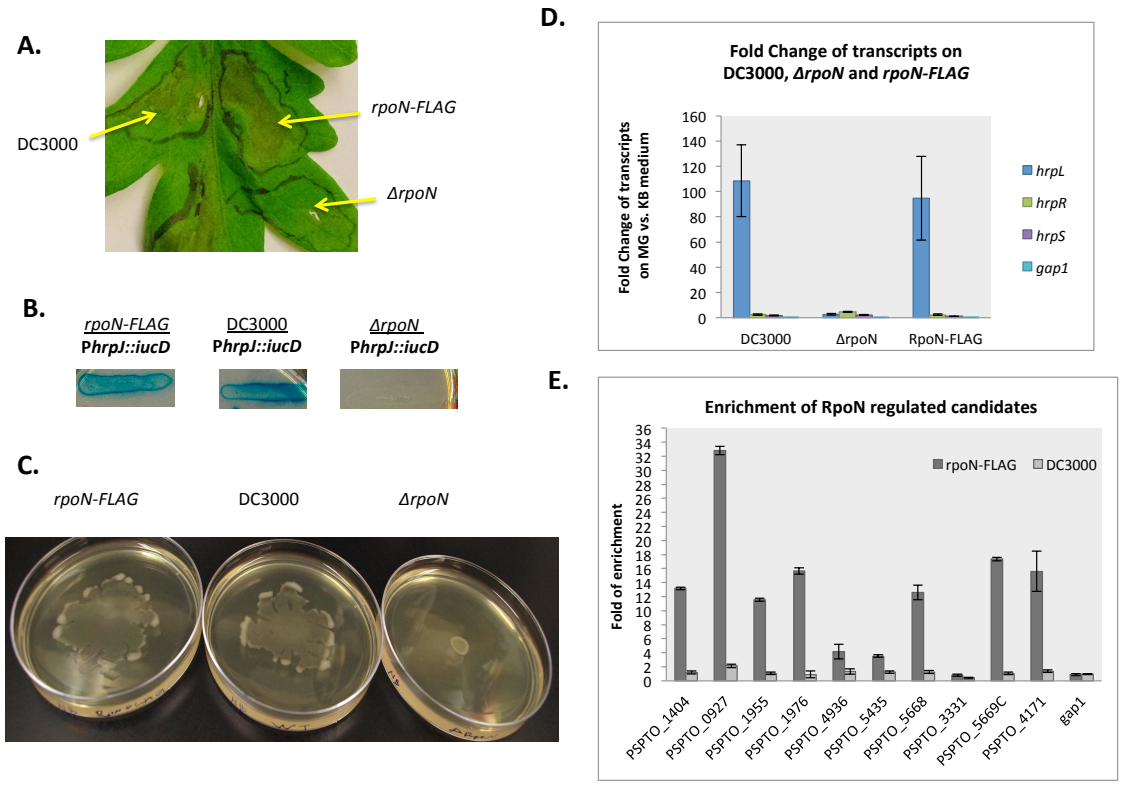


Figure 3. 3 RpoN-FLAG protein fusion is functional.

(A) *rpoN-FLAG* and WT DC3000 strains evoke the hypersensitive response while it is abolished in the *ΔrpoN* strain. Bacteria were infiltrated via a blunt syringe into three independent tomato plants at $3x10^7$ CFU/ml. Photos were taken after 2 days. Symptoms are identical among three replicates. (B) Expression of a plasmid-based (pBS181) (see Table 3.1) β-glucuronidase (GUS) reporter driven by a *hrp* promoter (*hrpJ*) in *rpoN-FLAG*, WT DC3000 and *ΔrpoN* backgrounds on HMM agar (*hrp*-minimal medium for induction of *hrp* promoters). All plates contain appropriate antibiotics to maintain plasmids and X-Gluc, a substrate of GUS. Pictures were taken 4 days after inoculation (C) Swarming activity of *rpoN-FLAG*, WT DC3000 and *ΔrpoN* strains. Swarming areas are similar among 3 replicates. Pictures were taken after 31 hours. (D) Relative transcript levels for *hrpL* (RpoN target), *hrpR*, *hrpS* and *gap1* (non RpoN targets) 1.5 hours after medium shift from KB to MG (supplemented with ferric iron at 50 μM final concentration) in *rpoN-FLAG*, WT DC3000 and *ΔrpoN* strains. *rpoN-FLAG* retains the ability to induce *hrpL* transcription similar to WT DC3000 while *ΔrpoN* does not. Error bars represent the standard deviation of two technical replicates. (E) Enrichment of potential RpoN targets. *rpoN-FLAG* samples are enriched at most potential targets while WT DC3000 samples show no detectable enrichment. Error bars represent the standard deviation of two technical replicates.

Interestingly, the KB and MG ChIP-Seq profiles are nearly identical, based on the observation that Genetrack (a software package designed for analyzing and visualizing ChIP-Seq data [153]) initially identified the same peaks in these two samples. The two profiles were therefore additively combined and considered as a single data set in the analyses below. An example is shown in the region of the RpoN promoter upstream from *hrpL* (Figure 3.4). This result is consistent with the model proposed by Ramos [198] in which RpoN, as part of the RNA polymerase holoenzyme, can bind stably to DNA without initiating transcription. In the case of *hrpL*, HrpR/S is not active in KB due to degradation of HrpR by Lon protease and binding of HrpV to HrpS which blocks the binding site of HrpS to HrpR as demonstrated in *P. syringae* pv. *phaseolicola* [223].

We identified 226 enriched regions using Genetrack and the combined ChIP-Seq data sets. MEME [154] was then used to search for over-represented motifs in the regions containing enriched sequence reads. The single motif that was recovered (Figure 3.5A), contributed by 214 of the 226 input regions, resembles the canonical RpoN motif [103] with highly conserved "GG" and "GC" residues separated by 10 nucleotides. The complete list of enriched regions and RpoN motifs are shown in Table 3.4. As expected, RpoN binding and an accompanying motif are observed upstream of *hrpL* (PSPTO_1404*)*. The majority of motifs is in intergenic regions and can be linked to plausible transcription targets. However, about a third of RpoN motifs are found in intragenic regions and a small number of motifs are oriented in an antisense direction (Figure 3.5B). RpoN binding in intragenic regions has been observed by others but the function in gene regulation is unknown [108].

Table 3. 3 Sequence reads and mapping statistics of ChIP-Seq samples for RpoN-FLAG

	Total reads	Reads aligned uniquely to the chromosome
ChIP-Seq on KB	18,340,472.00	13,331,983.00
ChIP-Seq on MG	18,280,000.00	12,874,564.00

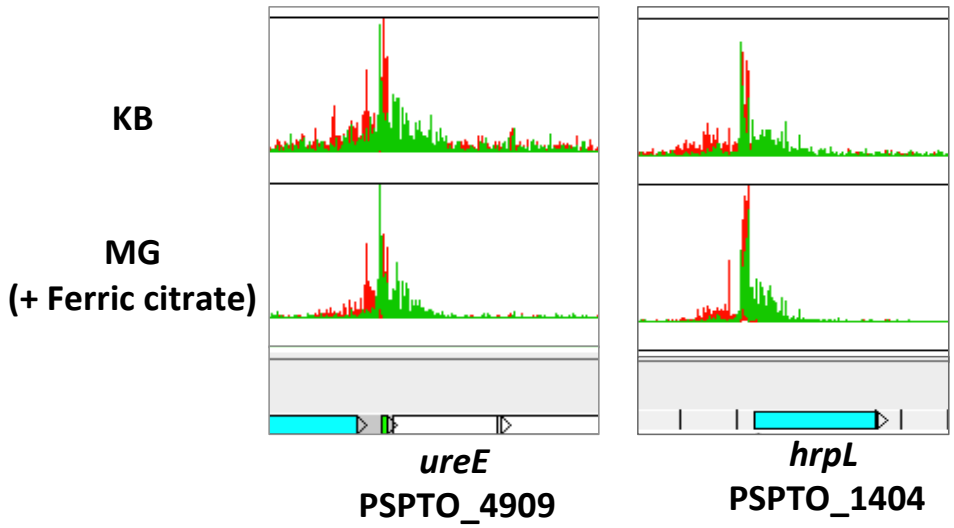


Figure 3. 4 ChIP-Seq profiles of *rpoN-FLAG* on KB and MG medium.

The ChIP-Seq profiles are identical on KB and MG (supplemented with Ferric citrate 50μM final concentration). Two genes are selected to represent the data. Red lines are for sequence reads mapping to the positive strand of chromosome, while green lines are for sequence reads mapping to the negative strand of the chromosome.

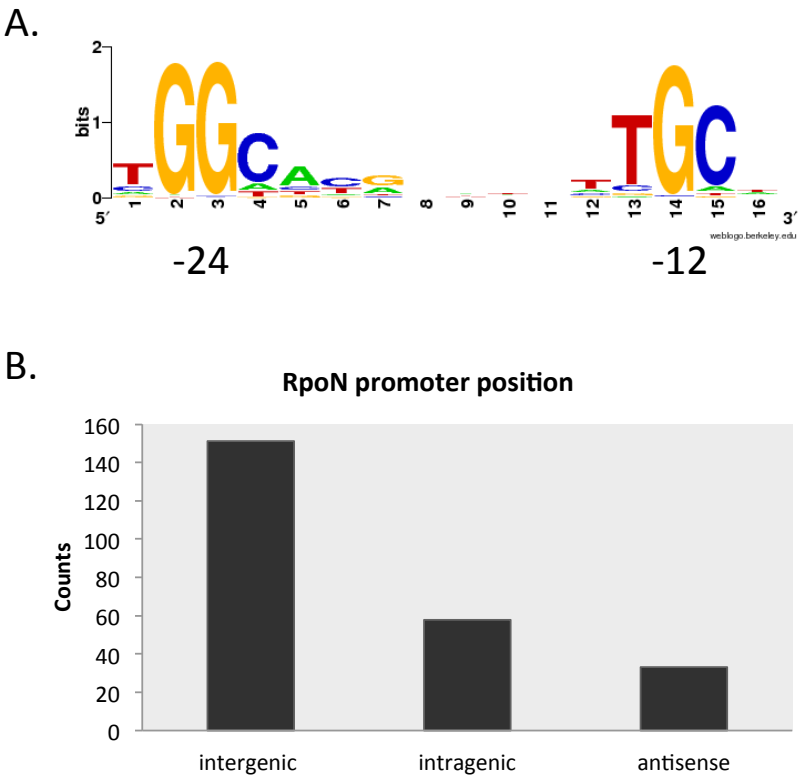


Figure 3. 5 RpoN motifs and their relative position in *Pseudomonas syringae* pv. *tomato* DC3000 genomes

(A) Motif logo for over-represented sequences associated with RpoN binding regions. (B). Positions of RpoN promoter motifs in DC3000 genomes. The 'intergenic' class contains motifs with positions consistent with normal promoter function; the 'intragenic' class contains motifs located within a coding region and oriented in the same direction as the gene; the 'antisense' class contains motifs located in coding regions but oriented in the opposite direction relative to the gene.

Table 3. 4 RpoN motifs identified in ChIP-Seq

"No": numerical label of RpoN binding sites. Only the first RpoN motif coordinate is designated with a number if more than one motif was found for a certain RpoN binding site.
"Motif coord start": position of 5' end of RpoN motif. For RpoN binding sites without associated motifs, the motif coordinate is the position of RpoN binding site. RpoN binding site coordinates have length of 20 bps centered at the highest enriched ChIP-Seq read position.
"Motif coord stop": position of 3' end of RpoN motif.
"Motif sequence": Only RpoN binding sites with associated RpoN motifs are shown with a motif sequence.
"st": strand on which RpoN motif was found. "+" represents the positive strand while "-" represents the negative strand on the main chromosome.
"Gene downstream of RpoN motif": First gene downstream of RpoN motif within 1000 bps.
"antisense": RpoN motif located in the coding region but on the opposite direction with the annotated gene; "intragenic": RpoN motif located in the coding region on the same direction with the annotated gene; "unknown": RpoN motif located in the intergenic region not in vicinity of any genes.
"Function": annotated function of the first gene downstream of RpoN motif within 1000 bps.

No	motif coord start	motif coord stop	motif sequence	st	Gene downstream of RpoN motif	Function
1	33323	33336	CGGCACGATATAAGCT	-	PSPTO_0026	hypothetical protein
2	38411	38424	AGGGTGGAAACCTGCA	+	PSPTO_0031	Ser/Thr protein phosphatase family protein
3	50449	50465	TGGTAGACGCTTTGCT	+	PSPTO_0038	conserved domain protein
4	60329	60342	TAGCCCGAATTTTGCT	-	unknown	
5	65877	65890	TGGCAATTTTCGTGCG	-	antisense	PSPTO_0047
6	68983	68996	TGCAAACGATGAGCCA	-	antisense	PSPTO_0049
7	90323	90343				
8	97753	97769	TGGCTTGCCGCTTGAT	+	PSPTO_0074	rpoZ
9	129940	129953	TGGCATGTTTATTGAG	-	PSPTO_0110	metallo-beta-lactamase superfamily protein
10	157478	157491	AGGCCAGGTTCTTGCT	-	PSPTO_0134	algQ
11	193857	193870	TGGCACGGATCTTGAA	+	PSPTO_0172	hypothetical protein
12	212374	212389	TGGTTTCAGACTTGCA	-	PSPTO_0188	
	212470	212483	TGGCCACCTACCCGCC	+	PSPTO_0189	nitrilase, putative
13	237715	237728	TGGCACCTTTGATGCA	-	PSPTO_0211	L-lysine 6-monooxygenase
	237751	237766	GGGCAAAGACGGTGCA	-	PSPTO_0212	transcriptional activator LysR
14	240131	240144	TGGCAAGCTTTCTGCT	+	PSPTO_0217	nitrogen regulatory protein P-II
15	261036	261049	TAGGCACGATTTTCTC	+	PSPTO_0240	prevent-host-death family protein
16	283323	283336	TGGCACGAGTATTGAA	+	PSPTO_0260	conserved hypothetical protein
17	297903	297916	TGGCACAGCGGTTGCA	-	PSPTO_0271	tonB protein, putative
18	298667	298687				
19	309405	309418	GGGCAATTTTCATGCT	+	unknown	
20	315929	315949				
21	362399	362412	TGGCATGCAGTCTGCA	-	PSPTO_0332	hypothetical protein

22	367376	367389	AGGAATATTTTATGCA	-	PSPTO_0337	prevent-host-death family protein
23	387338	387351	TGGTTTGTTTTTTGCA	-	PSPTO_0354	hypothetical protein
24	391726	391739	TTGGCATGAAGGTCGC	+	antisense	PSPTO_0359
25	392565	392578	TGGCAGAATACTCGCG	-	PSPTO_0359	glutamine synthetase, type I
	392601	392614	TGGCATGCAAATTGCT	-	PSPTO_0359	glutamine synthetase, type I
26	456785	456798	TGGCTCAAAGCTTGCA	-	PSPTO_0409	TldD/PmbA family protein
27	510635	510648	CGGCATGAGATTTGCG	+	PSPTO_0465	transcriptional regulator, AraC family
28	549516	549529	TGGCAAGGTCTATGCC	-	PSPTO_0501	hopU1
29	563429	563444	TGGACGCCCTCATGCC	-	PSPTO_0514	coenzyme PQQ synthesis protein F
	563503	563516	TGGCATGATAGTTGAG	+	PSPTO_0515	ISPsy8, transposase OrfA
30	590726	590739	CGGCTTGTTTTTTGCC	-	PSPTO_0537	RNA polymerase sigma-70 factor
31	770347	770360	CGGCACAGACATTGCT	+	unknown	
32	795434	795447	CGGCACGAAAGGTGCT	+	PSPTO_0748	RIO1/ZK632.3/MJ0444 family protein
33	819519	819532	TGGCCAGGATTTTGCT	-	PSPTO_0769	ABC transporter, ATP-binding protein
34	900107	900120	TGGTCGCGATATTGGA	+	antisense	PSPTO_0830
35	967267	967280	AGGAAGTCCACATGCA	+	PSPTO_0890	sucrose porin precursor
36	1007145	1007158	TGGCACGACCTATGCT	+	PSPTO_0927	type IV pilus biogenesis protein
37	1046962	1046975	TGGCACAGCTTCTGCT	-	PSPTO_5668	scRNA
38	1109632	1109645	GGGATTGCCCATGGCC	+	PSPTO_1015	integrase/recombinase XerC, putative
39	1164033	1164046	CGGTCCGGTTCTTGCT	+	PSPTO_1066	methyl-accepting chemotaxis protein
40	1178386	1178399	GGGCATGGACCTTGCC	-	intragenic	PSPTO_1074
41	1271390	1271403	TGGCCCGCTTAATGCA	-	PSPTO_1157	bacterial luciferase family protein
42	1272471	1272486	TGGCATGGATGCTGCT	+	PSPTO_1159	bmp family protein
43	1290373	1290386	TGGCATTGCGCTTGCT	-	PSPTO_1174	glycosyl transferase, group 1 family protein
44	1297436	1297473	GGAATAATCCTTGCTT	-	PSPTO_1184	conserved hypothetical protein
45	1297926	1297939	GGGCATGAATCGTGAC	+	PSPTO_1186	flavodoxin

46	1298361	1298374	TGGCCTGAAGCTTGCA	-	PSPTO_1185	protein of unknown function
47	1317048	1317061	TGGCCCGAACCATGCA	-	antisense	PSPTO_1203
48	1337773	1337786	CGGCACTGCCAAAGCG	-	unknown	
49	1345991	1346011				
50	1377906	1377919	TGGTACGGGCTTTGCT	+	PSPTO_1255	amino acid ABC transporter, periplasmic amino acid-binding protein
51	1402734	1402750	TGGCCGTAAAATTGCT	-	PSPTO_1277	gcvH-2
52	1433043	1433063	TGGCACGGTGCCTGCA	+	PSPTO_1304	conserved protein of unknown function
53	1452129	1452142	TGGCGCACAACTTGCT	-	PSPTO_1320	conserved protein of unknown function
54	1466903	1466916	TGGCACAGAACGTGAT	-	PSPTO_1335	cyclohexadienyl dehydratase, putative
55	1472235	1472248	TGGCTCAAGAATTGCT	-	intragenic	PSPTO_1341
56	1500192	1500205	TGGCCCATTAAGTGCA	-	PSPTO_1361	amidase family protein
57	1518254	1518267	CGGCTCGGTTTATGCA	-	intragenic	PSPTO_1377
58	1542788	1542801	TGGCATGGTTATCGCT	+	PSPTO_1404	RNA polymerase sigma factor HrpL
59	1610779	1610792	CGGACATGGGCTCGCC	+	PSPTO_1466	conserved protein of unknown function
60	1675214	1675227	CGGCACGGTCCTTGCG	+	antisense	PSPTO_1520
61	1681766	1681779	TGGCATAAGTTACGCA	+	intragenic	PSPTO_1524
62	1697701	1697714	CGGCATGTACTTTGCT	+	PSPTO_1539	phosphatidate cytidylyltransferase
63	1734717	1734730	AGGCACAGACAATGCA	-	PSPTO_1573	conserved hypothetical protein
64	1778764	1778777	TGGCACAGCTTCTGCT	-	PSPTO_5669	scRNA
65	1819765	1819778	TGGTATGACATATGCT	+	unknown	
66	1843497	1843510	TGGCATGTCGTTTGCC	+	PSPTO_1672	DNA-binding response regulator
67	1851958	1851971	TGGCACAGCTATTGCT	+	PSPTO_1682	C4-dicarboxylate transport protein
68	1874971	1874984	TGGACGTCTTTTTGCA	+	PSPTO_1705	NLP/P60 family protein"
69	1906285	1906298	GGGCGTGGATTTTGCA	+	PSPTO_1738	conserved protein of unknown function
70	1934922	1934935	GGGCATGAAAACTGCG	+	PSPTO_1764	conserved hypothetical protein
71	1994674	1994687	TGGCTTTTCAGGCGCT	+	PSPTO_1826	arginine/ornithine ABC transporter, periplasmic arginine/ornithine-binding protein
	1994761	1994774	AGGCCGATGACCCGCC	+	PSPTO_1826	arginine/ornithine ABC

						transporter, periplasmic arginine/ornithine-binding protein
72	2047292	2047305	TGGCACAGCTTTTGCG	-	PSPTO_1871	cytosine/purine/uracil/thiamine/allantoin permease family protein
73	2093597	2093613	TGGTTCGATTATTGCT	+	PSPTO_1913	conserved protein of unknown function
74	2103374	2103387	TGGAGGGTTCTCTGCT	+	PSPTO_1921	glutamine synthetase
	2103625	2103638	TGGCGGCCTGGACGCT	+	PSPTO_1921	glutamine synthetase
75	2109153	2109169	CGGCACGGGCTTTGCT	-	PSPTO_1926	flagellar protein, putative
76	2119223	2119260	TGGCACGCCGATTGCT	+	PSPTO_1933	flagellar basal-body rod protein FlgB
77	2120018	2120034	TGGTTCAAAAATTGCT	+	PSPTO_1939	flagellar basal-body rod protein FlgF
78	2113909	2113922	TGGCTCGCGTCTTGCA	+	PSPTO_1940	flgG
79	2141400	2141413	CGGCACGAGGATTGCT	+	PSPTO_1955	sensor histidine kinase FleS
80	2144262	2144275	TGGCACCGTTGTTGCT	+	PSPTO_1957	flagellar hook-basal body complex protein FliE
81	2152639	2152652	TGGCCCACATATTGCT	+	PSPTO_1966	flagellar hook-length control protein FliK
82	2154613	2154626	TGGCATAACACTTGCT	+	PSPTO_1968	flagellar protein FliL, putative
83	2160407	2160420	TGGACAGCTTTTTGCA	+	PSPTO_1976	flagellar biosynthesis protein FlhA
84	2162329	2162342	TGGAACAGATATTGCT	+	PSPTO_1977	flagellar biosynthesis protein FlhF
85	2172090	2172103	CGGCACGGCCAATGCG	+	PSPTO_1986	ParA family protein
	2172132	2172145	TGGCACACAAACTGCA	+	PSPTO_1986	ParA family protein
	2231826	2231841	TGGCGGCTCCATTGCT	+	PSPTO_2050	lipoprotein, putative
	2231858	2231873	TGGAAGTACTATTGCA	+	PSPTO_2050	lipoprotein, putative
86	2231871	2231884	TGGCACAGACACTGCA	-	PSPTO_2049	conserved hypothetical protein
87	2239492	2239505	TGGCACGGGCTTTGCG	+	PSPTO_2059	conserved protein of unknown function
88	2263816	2263831	AGGAGCTTAGCGTGCA	+	PSPTO_2093	lysozyme, putative
	2263869	2263882	TGGCGTGATTTATGCC	-	antisense	PSPTO_2092
89	2507808	2507825	CGGCATGACACTTGCT	-	unknown	
90	2548845	2548858	TGGCACAAGCGCTGCA	-	PSPTO_2302	assimilatory nitrate reductase electron transfer subunit domain protein
91	2552066	2552079	TGGCCCATCTCTTGCT	-	PSPTO_2304	nitrate transporter
92	2552748	2552761	CGGCATGCAGGTTGCA	-	unknown	
93	2556054	2556067	TGGCATATGCCTTGCT	-	PSPTO_2308	hypothetical protein
94	2564599	2564612	CGGCTCGTTTCATGCG	-	PSPTO_2314	D-alanyl-D-alanine carboxypeptidase
95	2579313	2579326	GGGCATGAGAAATGCG	-	PSPTO_2324	ISPssy, transposase
96	2629058	2629071	TGGCACGAAACCCGCT	-	antisense	PSPTO_2378

97	2667265	2667278	TGGCACTGTGCTTGCT	-	PSPTO_2418	branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein
98	2691789	2691802	TGGAACACCTTTTGCA	+	PSPTO_2439	acyl-CoA dehydrogenase family protein
99	2733321	2733334	AGGCACGGTGATTGCT	+	PSPTO_2475	methyl-accepting chemotaxis protein
100	2804455	2804468	TGGCATGATATTTGTT	-	PSPTO_2539	secreted protein Hcp
101	2807147	2807160	CGGCACGCCATTTGCC	+	PSPTO_2543	conserved hypothetical protein
102	2821365	2821378	GGGCATGTATTGTGCA	+	PSPTO_2555	phosphonates ABC transporter, ATP-binding protein
103	2851488	2851505	TGGCAAGCGCTTTGCA	+	PSPTO_2582	glutamine synthetase
104	2860315	2860328	TGGCACGGTTGCTGCT	+	PSPTO_2590	bacterial luciferase family protein
105	2991616	2991629	TGGCAGATCTCTTGAC	-	PSPTO_2697	oxidoreductase, zinc-binding protein
	2991634	2991647	AGGTATGAACTTTGCA	+	PSPTO_2698	protein of unknown function
106	2994750	2994763	TGGCACCAGACTTGCT	+	PSPTO_2700	protein of unknown function
107	3029252	3029265	TGGCCCGTCGCTTGCT	+	PSPTO_2731	periplasmic sugar-binding domain protein
	3029283	3029295	TGGTCAATTCTGCGCG	+	PSPTO_2731	periplasmic sugar-binding domain protein
108	3047275	3047291	TGGCACGGGTCCTGCT	-	antisense	PSPTO_2741
109	3094326	3094339	TGGCACACTTTCTGCT	+	PSPTO_2775	amino acid ABC transporter, periplasmic amino acid-binding protein
110	3104157	3104170	AGGCACATTTCGTGCT	+	PSPTO_2785	putrescine ABC transporter, periplasmic putrescine-binding protein
111	3108041	3108056	TGGCACAGAGCGTGCA	-	PSPTO_2788	ABC transporter, periplasmic substrate-binding protein
112	3116330	3116343	AGGCCCGCAACTTGCA	+	PSPTO_2796	conserved hypothetical protein
113	3119710	3119723	TGGCGATGATTTGGCC	+	PSPTO_2797	hypothetical protein
114	3133633	3133646	TGGAATGATTTTTGAA	+	PSPTO_2809	conserved hypothetical protein
115	3243860	3243873	TGGCACACCCCTTGCA	+	intragenic	PSPTO_2880
116	3248710	3248723	TGGTAAAATCTTGGCA	+	unknown	
117	3253403	3253416	TGGCTTGCCGATTGCA	+	PSPTO_2893	PAP2 superfamily protein
118	3315954	3315967	CGGCACAGGCTTTGCT	-	PSPTO_2950	benzaldehyde dehydrogenase (NAD+)
	3315968	3315983	GGGATTGCAACCTGCT	+	PSPTO_2951	sigma-54 dependent transcriptional regulato
119	3350548	3350564	TGGCACAGTTTCTGCT	-	PSPTO_2980	oxidoreductase, 2OG-Fe(II) oxygenase family
120	3351387	3351403	AGGCATAACCGAGGTA	-	PSPTO_2981	isochorismatase family protein
121	3357087	3357100	CGGCACGTACTGTGCA	+	PSPTO_2985	acyl-CoA dehydrogenase

						family protein
122	3363667	3363680	GGGCACAAAGCATGCA	-	PSPTO_2992	acyl-CoA dehydrogenase family protein
	3363699	3363712	TGGCACGGTCGCTGCT	+	intragenic	PSPTO_2991
123	3369632	3369645	GGGCATGACCTGTGCA	-	PSPTO_2996	ribose ABC transporter, periplasmic ribose-binding protein, putative
124	3400324	3400337	CGGCATGGTTGATGCC	-	PSPTO_3023	excinuclease ABC, C subunit
125	3404447	3404460	CGGCATGCTCTATGCG	+	intragenic	PSPTO_3029
126	3427989	3428002	TGGCACAGGCTTTGCT	+	PSPTO_3047	myo-inositol 2-dehydrogenase, putative
	3428002	3428015	TGGTTTGAATATTGGC	+	PSPTO_3047	myo-inositol 2-dehydrogenase, putative
127	3518158	3518173	TGGCGCAAACCTTGCT	+	PSPTO_3129	conserved protein of unknown function
128	3525745	3525758	TGGCCCGGCTTGTGCA	-	PSPTO_3135	conserved hypothetical protein
129	3558702	3558715	TGGCATGCTTGCTGCT	+	PSPTO_3166	conserved domain protein
130	3598631	3598644	TGGCGTGATTGTTGCC	-	PSPTO_3204	conserved hypothetical protein
131	3611631	3611644	TGGCACAGGCCTTGCC	+	unknown	
132	3686217	3686230	TGGCACAACCTCTGCA	+	PSPTO_3262	nitrite reductase [NAD(P)H], large subunit
133	3717349	3717362	AGGCGAAGATAACGCT	-	unknown	
	3717198	3717211	AGGTGTGATTCCGGCG	-	unknown	
134	3828463	3828476	TGGCACGATTCGTGGA	-	PSPTO_3386	protein of unknown function
135	3894852	3894867	TGGCACACAAGCTGCT	+	PSPTO_3451	ssue
136	3917889	3917902	TGGCACGTATGATGCT	+	PSPTO_3471	monovalent cation:proton antiporter, putative
137	3941459	3941472	CGGCACGATAGTTGTA	-	unknown	
138	3941591	3941604	CGGCACGATAGTTGTA	+	unknown	
139	3969910	3969925	CGGGTATTGTTCAGCT	-	PSPTO_3517	fatty oxidation complex, alpha subunit
	3969947	3969960	AGGCTACGACGTCGCC	+	unknown	
140	3997613	3997633				
141	4077381	4077394	GGGCTCAAAAGACGGA	+	antisense	PSPTO_3616
142	4126336	4126349	TGGCACGAAATACGCA	+	PSPTO_3660	xanthine dehydrogenase, N-terminal subunit
143	4134412	4134425	AGGCAGGTTTCTTGCT	+	PSPTO_3666	polysaccharide deacetylase family protein
144	4139223	4139239	TGGCGCAGCTCTTGCT	+	PSPTO_3671	conserved hypothetical protein
145	4187789	4187803	CGGACAATTTCATGCC	+	intragenic	PSPTO_3709
146	4225510	4225523	TGGCACGAATGTCGCT	-	PSPTO_3740	ABC transporter, ATP-binding protein
147	4283344	4283367	AGGCATTCAGCCAGCG	+	unknown	
	4283471	4283486	TGGCATTCAGGTTGCA	-	PSPTO_3779	conserved hypothetical

						protein
148	4305918	4305931	TGGCATGAATTCTGCG	-	PSPTO_3801	major facilitator family transporter
149	4331292	4331305	TGGCGCGGAAGCTGCA	+	PSPTO_3821	lipoprotein, putative
150	4389842	4389855	TGGCATGGTTCGTGAT	-	PSPTO_3877	transcriptional regulator, GntR family
151	4389960	4389980				
152	4410301	4410321				
153	4418232	4418245	AGGCATGAAAAATGTC	+	PSPTO_3903	conserved protein of unknown function
154	4457330	4457350	TGGCCCGATCATTGCA	-	unknown	
155	4458185	4458198	AGGAGGCCATATCGCT	+	antisense	PSPTO_3949
	4458275	4458288	TGGCTCCCTTCTTGCT	+	antisense	PSPTO_3949
156	4492468	4492481	AGGCACTGTGCCTGCT	+	PSPTO_3987	porin D
157	4507211	4507224	GGGCATAAAAGGTGCC	+	PSPTO_3995	internal deletion…
158	4535319	4535334	CGGAGAGTAAATCGCA	-	PSPTO_4024	ferredoxin--NADP reductase
	4535389	4535402	CGGTGCTGCTTTCGCT	+	intragenic	PSPTO_4025
159	4544951	4544964	TGGCGTGATTTATGCC	+	antisense	PSPTO_4038
160	4563288	4563302	GGGATTCCGAGCTGCT	+	PSPTO_4058	mutS
161	4584300	4584313	GGGCATGGATATTGCT	+	PSPTO_4079	sensor histidine kinase/response regulator
162	4599142	4599155	TGGCACGATAGTTGAG	-	unknown	
163	4599324	4599339	TGGCACGATAGTTGAG	-	unknown	
	4599346	4599361	TGGCACGATAGTTGAG	-	unknown	
164	4620294	4620307	CGGCATGAATCTTGTG	+	PSPTO_4100	MaoC-like domain protein
165	4631277	4631290	TGGCATAGAGGGTGCT	+	PSPTO_4108	high-affinity branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein
166	4663129	4663166	TGGTTCGGAAATTGCT	+	PSPTO_4136	amino acid ABC transporter, periplasmic amino acid-binding protein
167	4679009	4679022	TGGCAAGTTTTATGTA	-	PSPTO_4151	DNA-binding response regulator, LuxR family
168	4684834	4684847	GGGCATGATGATTGCG	-	PSPTO_4156	sodium-type flagellar protein MotY, putative
169	4692085	4692098	TGGTGCAAGATCTGAA	+	PSPTO_4165	ABC transporter, ATP-binding protein
	4692053	4692066	CGGCTCGAAACCGGCT	-	antisense	PSPTO_4164
170	4699434	4699454	TGGCACGACTCATGCC	+	PSPTO_4171	amino acid ABC transporter, periplasmic amino acid-binding protein
171	4736896	4736909	TGGCATGGAATCTGCC	-	PSPTO_4204	amidase family protein
172	4771306	4771319	TGGCACAACTCTCGCT	+	PSPTO_4238	ABC transporter, periplasmic substrate-binding protein
173	4772455	4772468	TTGCCCACGGCTTGCT	+	PSPTO_4239	ABC transporter, permease

						protein
	4772501	4772514	CGGCATGACATTTGCC	+	PSPTO_4239	ABC transporter, permease protein
174	4842764	4842779	TGGCACAGCTCGTGCT	+	PSPTO_4296	metabolite-proton symporter, putative
175	4811660	4811680				
176	4842764	4842779				
177	4929604	4929619	CGGCATTTTTTATGCC	+	PSPTO_4367	lipoprotein, putative
	4929606	4929619	AGGCATAAAAAATGCC	-	antisense	PSPTO_4366
178	4969283	4969303				
179	5017144	5017157	AGGCATATAATTTGCT	+	PSPTO_4453	RNA polymerase sigma-54 factor
180	5050570	5050583	TGGCACACCGATTGCG	-	PSPTO_4484	conserved protein of unknown function
181	5063867	5063880	CGGCATGAAAAGTGCG	+	antisense	PSPTO_4495
182	5066863	5066883				
183	5114725	5114738	CGGAATGAAACTTGCC	-	PSPTO_4528	transcriptional regulator, AraC family
184	5118897	5118910	TGGCCCGCGAAATGCT	-	intragenic	PSPTO_4531
	5126993	5127008	TGGCACTGTCCTTGCT	+	PSPTO_4539	transcriptional regulator, LuxR family autoinducer-regulated
185	5164806	5164821	CGGCAGGTCTGGTGCT	.	PSPTO_4569	iron-uptake factor
186	5164894	5164907	CGGCGCAATTAATGCA	+	antisense	PSPTO_4570
187	5208002	5208017	CGGGTCAGTGCAGGCA	-	PSPTO_4611	conserved domain protein
	5207982	5207995	TGGTTGAACTGCCGCA	+	antisense	PSPTO_4612
188	5208643	5208656	TGGTATCCAAGGCGCT	+	PSPTO_4613	C4-dicarboxylate transporter/malic acid transport protein
	5208765	5208778	TGGCGTAGGCTTAGCT	+	PSPTO_4613	C4-dicarboxylate transporter/malic acid transport protein
189	5225860	5225873	CGGAACGCTTCTTGCT	-	antisense	PSPTO_4627
190	5328782	5328795	TGGCGCGATAACTGCA	-	intragenic	PSPTO_5629
191	5329403	5329423	TGGAACAGAAAATGCA	-	PSPTO_5629	insertion sequence, putative
192	5376565	5376578	AGGCGAGCGACTGGCA	-	PSPTO_4745	ATP-dependent helicase HrpB, putative
	5376722	5376737	TGGATCTCTTTGAGCG	+	PSPTO_5632	hypothetical protein"
193	5425809	5425824	AGGTTTTGCTAAAGCC	+	intragenic	PSPTO_4786
194	5437013	5437026	GGGCATGGTTGATGCG	+	antisense	PSPTO_4796
195	5438513	5438526	TTTGCACGTGTTTTGC	-	PSPTO_4797	protein of unknown function
196	5445807	5445820	TGGAACGAACGTTGCT	+	PSPTO_4807	metal ion transporter, putative
197	5535074	5535089	TGGCCCAGACCTTGCT	+	PSPTO_4885	branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein

198	5558351	5558371	TGGCAAGCCCCTTGCT	+	PSPTO_4909	ureE
199	5568952	5568965	TGGCACGCGCCTTGCT	-	PSPTO_4919	high affinity branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein
200	5580076	5580089	TGGCATAAGTCTTGCG	-	PSPTO_4926	transglutaminase-like superfamily domain protein
201	5588639	5588659				
202	5595397	5595410	CGGCACGGCGCTTGCT	-	PSPTO_4936	methyl-accepting chemotaxis protein
203	5598123	5598136	GGGCACGACAAATGCC	+	antisense	PSPTO_4939
	5598123	5598138	GGGCATTTGTCGTGCC	-	PSPTO_4938	ATP phosphoribosyltransferase regulatory subunit, putative
204	5636792	5636805	AGGCATGGATACTGCG	-	antisense	PSPTO_4969
205	5717647	5717664	TGGCACGCGGACTGCA	+	PSPTO_5023	sodium/hydrogen exchanger family protein
206	5737978	5737978				
207	5743662	5743675	CGGTGACCGGCTTGCC	+	PSPTO_5044	NLP/P60 family protein
208	5841131	5841168	TGGCGCTCGATTTGCA	-	PSPTO_5132	pilM
209	5853626	5853639	TGGCAAGTTTCCTGCG	+	PSPTO_5143	contain authentic mutation
210	5858972	5858985	TGGCACGATAGTTGAA	+	antisense	
211	5877841	5877854	TGGAACGATCGATGCA	+	PSPTO_5164	proline iminopeptidase
	5877853	5877866	TGGCAAAACGCTTGCA	-	PSPTO_5164	proline iminopeptidase
212	5948551	5948564	TGGCCTTTAAGTTGCT	-	PSPTO_5224	conserved protein of unknown function
213	5964492	5964505	TGGCATGCGCTGTGCA	+	PSPTO_5245	amino acid ABC transporter, periplasmic amino acid-binding protein
214	6035685	6035698	CGGCACCTTATTTGCA	-	PSPTO_5307	putrescine ABC transporter, periplasmic putrescine-binding protein
215	6051673	6051686	TGGCACGCACTCTGCA	+	antisense	PSPTO_5322
216	6089515	6089528	TGGCGTAATTCTTGCG	-	PSPTO_5356	gamma-aminobutyrate permease
217	6109320	6109333	TGGCATGCTCATTGCT	+	PSPTO_5377	membrane protein, putative
218	6122972	6122985	TGGTCCCGATCTTGCT	-	PSPTO_5392	protein of unknown function
219	6181011	6181048	TACATGATTTTTTTGT	+	PSPTO_5435	hcp-2; secreted protein Hcp
220	6188865	6188878	TGGCAGAAGATATGCA	+	PSPTO_5439	
221	6189471	6189491				
222	6237634	6237647	TGGTACAAACCGTGCG	+	PSPTO_5478	phosphate regulon sensor protein phoR
223	6256023	6256036	AGGCACCAAAATTGCG	+	PSPTO_5489	cytosolic long-chain acyl-CoA thioester hydrolase family protein
224	6279093	6279106	AGGCCTTATACATGCG	+	PSPTO_5512	transcriptional regulator, LysR family

225	6361366	6361379	TGGCTCATTGAATGCG	-	unknown	
226	6388042	6388055	CGGCACAAATTATGCC	-	PSPTO_5606	ATP synthase F0, I subunit
	6388041	6388055	TGGCATAATTTGTGC	+	antisense of PSPTO_5606	

Although many of the RpoN motifs are solitary, 30 of the enriched regions are associated with two motifs in close proximity, and one region may have up to three RpoN motifs. The candidate promoters in these 31 regions are arranged in different configurations relative to one another (Figure 3.6). RpoN motifs in tandem orientation are found upstream of genes such as PSPTO_0359 (glutamine synthetase, type I), PSPTO_1921 (glutamine synthetase), PSPTO_1986 (ParA family protein), and PSPTO_3047 (myo-inositol 2-dehydrogenase, putative). RpoN motifs in convergent orientation appear upstream of genes such as PSPTO_2093 (lysozyme, putative), and PSPTO_4611 (conserved domain protein). Finally, divergent RpoN motifs occur upstream of genes such as PSPTO_3517 (fatty oxidation complex, alpha subunit), PSPTO_3779 (conserved hypothetical protein), and PSPTO_4367 (lipoprotein, putative) (divergent and overlapped).

In cases where more than one RpoN motif is present, it is sometimes difficult to determine whether only one, or possibly all, is associated with RpoN binding. For example, RpoN binding is detected in the region between PSPTO_5606 and PSPTO_5607. The RpoN motif at coordinate (6388042..6388055) on the complementary strand is upstream from PSPTO_5606, also known as *aptI*, and is also directly upstream from one of our captured 5'-ends. However, on the other strand directly opposite to this motif is another with sequence 5'-tggcataatttgtgc-3'. This motif is directly upstream from a captured 5'-end reported by Filiatrault *et. al.* [62]. It is possible that both motifs are genuine RpoN promoters but support transcription under different conditions. Note that Filiatrault *et. al*. [62] also identified an RpoD promoter in this region, oriented toward PSPTO_5606 at coordinate (6388066..6388097) on the complementary strand.

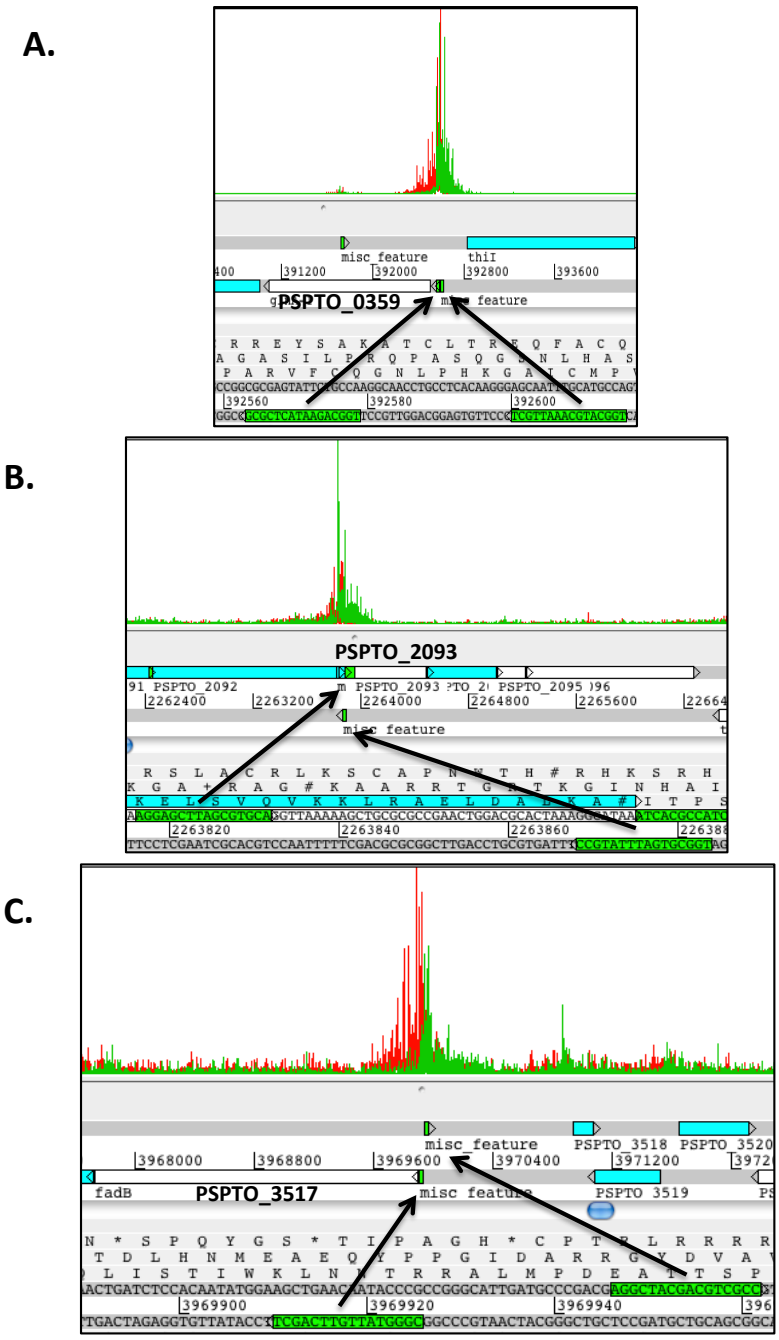


Figure 3. 6 Regions containing multiple RpoN motifs.

Artermis screenshots of genomic regions exhibiting RpoN binding (ChIP-Seq peaks) and multiple associated RpoN motifs. Examples of (A) tandem promoters; (B) convergent promoters; (C) divergent promoters. The ChIP-Seq profile is at the top of each panel. RpoN motifs are in green with corresponding sequences shown at the bottom of the panel. Coding regions and a more compact representation of the RpoN motifs are shown in the panel center.

The genes listed in Table 3.4 represent several metabolic "themes" that are summarized in Figure 3.7. The themes include flagella biosynthesis, nitrogen metabolism (glutamine synthetase, proline dehydrogenase), transport and binding proteins (ABC transporters), central metabolism (lipase, and enzymes in CoA pathway), putative regulatory proteins and small noncoding RNAs. This classification scheme is not definitive because proteins can be listed in multiple pathways (for example, *hrpL* in 'cellular processes' or 'transcription'), although only one classification is used per protein in the figure. Interestingly, genes involved in transport and binding (21%) appear to be enriched relative to their frequency in the genome as a whole (15%) summarized by Bao et. al. [224]. In contrast, genes associated with mobile DNA are more common in DC3000 (around 8%) than in our data (0%). This suggests that genes regulated by RpoN are not often horizontally transferred (by mobile DNA elements), and RpoN may have a greater responsibility to regulate the transport of nutrients and/or toxins in and out of the bacterial cells.

It is unlikely that all the genes listed in Table 3.4 are ever expressed at the same time. The most obvious example of differential expression is *hrpL*, which is induced under specific growth conditions. The ChIP-Seq data clearly show that RpoN binding can be detected upstream from *hrpL* in both *hrp*-inducing and non-inducing media (Figure 3.4). In contrast, transcription is detected (by 5'-capture) only in inducing medium (see below). As noted earlier, RpoN appears to bind stably to its promoters in a 'closed complex' and does not initiate transcription until an appropriate enhancer is available [105, 106]. Enhancer binding proteins probably function by inducing specific subsets of the regulon at different times in response to changes in cell state or environmental signals. The DC3000 genome

contains 22 likely enhancer proteins based on domain and sequence homology with known enhancers [215], including HrpS, the enhancer associated with *hrpL* transcription. Experiments designed to probe transcriptional activity in regions where RpoN binds are presented below.

With RpoN binding activity demonstrated and likely RpoN promoter motifs identified, the DC3000 regulon of RpoN can be, at least partially, defined. Because an arbitrarily high cutoff value was used to select enriched regions in the Genetrack ChIP-Seq analysis, it is likely that additional regions could be recovered with less stringent criteria. However, if the current model for RpoN binding and transcription is correct, we would not expect repeated ChIP-Seq experiments, even in different culture media, to fundamentally change the results. Instead, we anticipate that the functional characterization of RpoN sub-regulons will require experiments that examine transcription as it occurs in response to various conditions.

Differential gene expression using RNA-Seq

RNA-Seq is now widely used to assess global transcription patterns in bacteria. More recent protocols provide strand-specific information and can be modified to capture the 5'-ends of mRNA molecules and identify candidate transcription start sites (see chapter 2). RNA-Seq was used to examine DC3000 transcription in cells grown under *hrp*-inducing and repressing conditions as described above. Three biological replicate samples were collected for each condition. Table 3.5 summarizes the reads and mapping statistics for the 12 libraries using data generated by the Illumina Genome Analyzer.

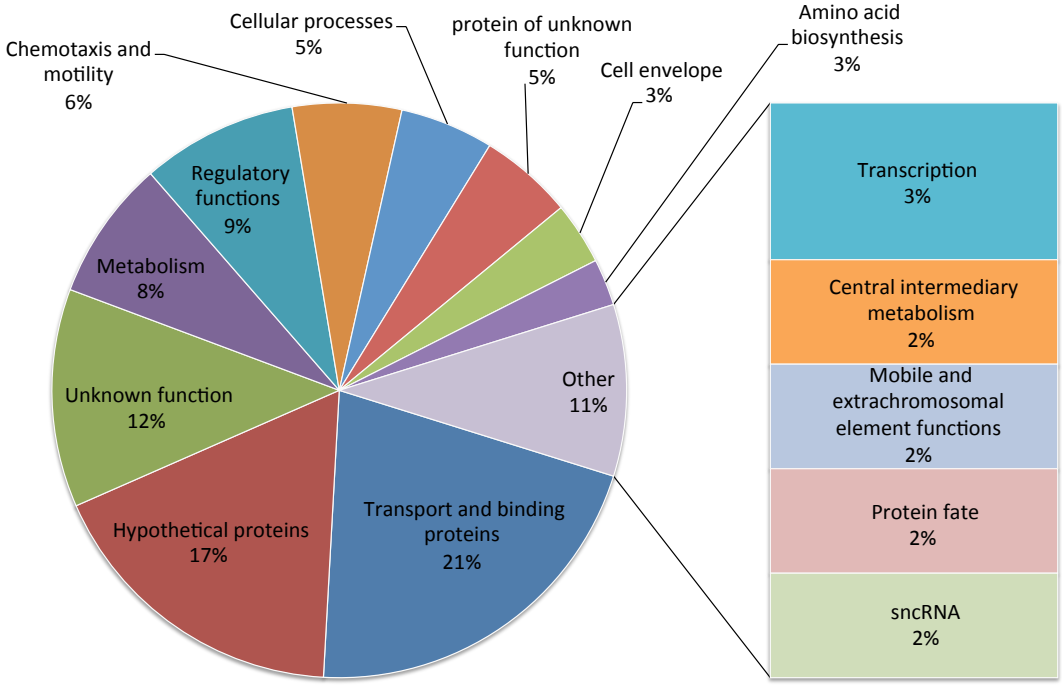


Figure 3. 7 Roles of RpoN regulon members

Functional categories are summarized based on the JCVI database [220].

Table 3. 5 Sequence reads and mapping statistics of RNA-Seq samples from *hrp*-inducing and non-inducing conditions

	Total reads	Reads aligned uniquely to the chromosome
RNA-Seq on KB replicate 1	21,643,516.00	9,317,820.00
RNA-Seq on KB replicate 2	23,102,825.00	10,528,434.00
RNA-Seq on KB replicate 3	7,696,972.00	3,778,067.00
RNA-Seq on MG replicate 1	19,873,668.00	13,000,850.00
RNA-Seq on MG replicate 2	21,965,580.00	6,723,420.00
RNA-Seq on MG replicate 3	28,318,651.00	10,767,956.00
5'capture on KB replicate 1	354,531.00	94,170.00
5' capture on KB replicate 2	579,484.00	131,434.00
5' capture on KB replicate 3	334,303.00	155,961.00
5' capture on MG replicate 1	452,884.00	261,874.00
5' capture on MG replicate 2	557,572.00	118,496.00
5' capture on MG replicate 3	420,225.00	148,420.00

Analysis of RNA-Seq data is reproducible

Before comparing RNA-Seq results between experimental conditions, the data sets were tested for reproducibility. Gene expression, as normalized read counts per gene summarized by reads per kilobase per million mapped reads (RPKM), was compared for all genes within and between biological replicates. The three biological replicates corresponding to each condition show high correlation ($R \geq 0.91$) (Figure 3.8A-D).

Reads from the two experimental conditions were compared using DE-Seq [221], a package available in the R statistical environment for the assessment of differential mRNA expression. Using DE-Seq, the comparison of RNA from *hrp* induced and repressed cells reveals 225 differentially expressed genes at a p-value cutoff 0.001 (Table 3.6). A graphical comparison between conditions shows that the majority of genes are expressed at similar levels except for the 225 genes called by DE-Seq (green, red and blue dots) (Figure 3.8E). As expected, all HrpL regulated genes are up-regulated in MG (red stars), while RpoN regulated genes can be up or down-regulated (blue stars and circles, respectively).

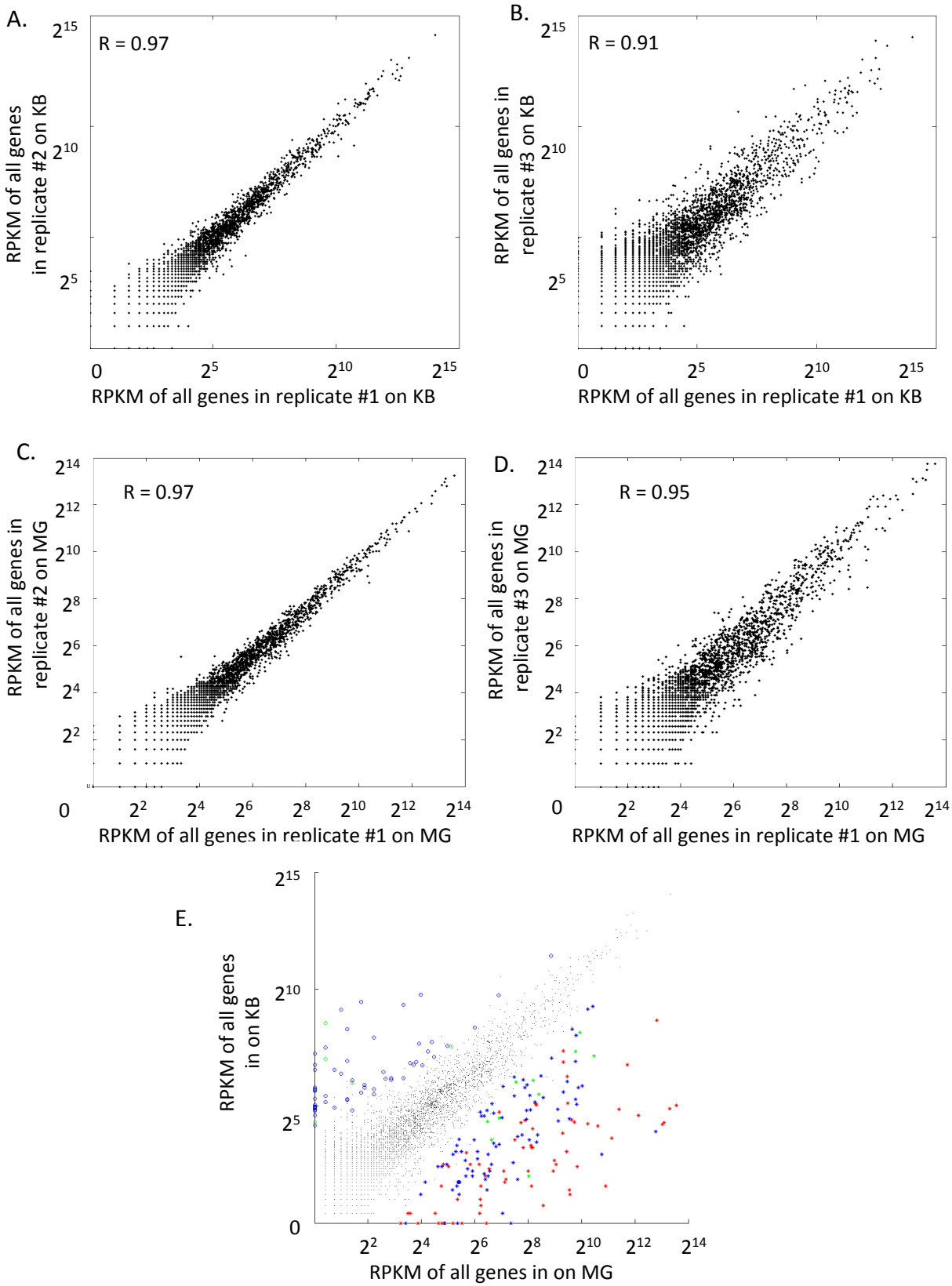

A.
R = 0.97
RPKM of all genes in replicate #2 on KB
RPKM of all genes in replicate #1 on KB
B.
R = 0.91
RPKM of all genes in replicate #3 on KB
RPKM of all genes in replicate #1 on KB
C.
R = 0.97
RPKM of all genes in replicate #2 on MG
RPKM of all genes in replicate #1 on MG
D.
R = 0.95
RPKM of all genes in replicate #3 on MG
RPKM of all genes in replicate #1 on MG
E.
RPKM of all genes in on KB
RPKM of all genes in on MG

Figure 3. 8 Variation in RNA-Seq.

Expression of all 5793 genes in each sample was computed as reads per kilo base per million reads mapped (RPKM). Gene expression was plotted for 2 of 3 replicates on KB (A, B), and on MG (C, D). Average gene expression in 3 replicates was plotted for KB versus MG (E). Shape indicates whether genes were upregulated (asterisk) or downregulated (circle) in the DE-Seq analysis. Color indicates whether genes are regulated by HrpL (red), by RpoN (green), or by an unknown sigma factor (blue). Regulators are inferred from ChIP-Seq and/or promoter search results, or published data (HrpL; chapter 2).

Table 3. 6 List of differentially expressed genes in *hrp*-inducing compared to non-inducing conditions

PSPTO	Function	Category roles from JCVI database
44	type III effector HopK1	Protein fate
61	type III effector HopY1	Protein fate
67	tonB protein	Transport and binding proteins
68	TonB system transport protein ExbD	Transport and binding proteins
69	TonB system transport protein ExbB	Transport and binding proteins
167	protein of unknown function	
307	conserved protein of unknown function	Hypothetical proteins
308	sulfate-binding protein	Transport and binding proteins
369	outer membrane porin, OprD family	Transport and binding proteins
370	MATE efflux family protein	Transport and binding proteins
371	indoleacetate-lysine ligase	Cellular processes
443	choline dehydrogenase	Cellular processes
500	acyl-CoA dehydrogenase family protein	Fatty acid and phospholipid metabolism
501	type III effector HopU1	Protein fate
503	type III chaperone protein ShcF	Cellular processes
524	peptidase, M20/M25/M40 family	Protein fate
536	sensory box/GGDEF domain/EAL domain protein	Regulatory functions
588	type III effector HopH1	Protein fate
589	type III effector HopC1	Protein fate
653	bacterioferritin	Transport and binding proteins
834	alcohol dehydrogenase, zinc-containing protein	Energy metabolism
835	ribD C-terminal domain protein	Unknown function
836	conserved domain protein	Unknown function
837	conserved protein of unknown function	Hypothetical proteins
852	type III helper protein HopAJ1	Protein fate
873	amidinotransferase family protein	Cellular processes
874	nikkomycin biosynthesis domain protein	Unknown function
875	conserved protein of unknown function	Cellular processes
876	type III effector HopD1	Protein fate
877	type III effector HopQ1-1	Cellular processes
883	type III effector HopR1	Protein fate

981	acetolactate synthase, large subunit, biosynthetic type	Amino acid biosynthesis
994	carbonic anhydrase	Central intermediary metabolism
1022	type III effector HopAM1-1	Protein fate
1206	TonB-dependent siderophore receptor, putative	Transport and binding proteins
1255	amino acid ABC transporter, periplasmic amino acid-binding protein	Transport and binding proteins
1257	amino acid ABC transporter, permease protein	Transport and binding proteins
1275	glycine cleavage system T protein	Energy metabolism
1276	glycine dehydrogenase	Energy metabolism
1277	glycine cleavage system H protein	Energy metabolism
1292	glucose ABC transporter, periplasmic glucose-binding protein, putative	Transport and binding proteins
1296	porin B	Transport and binding proteins
1334	methyl-accepting chemotaxis protein	Cellular processes
1369	type III chaperone protein ShcN	Cellular processes
1370	type III effector HopN1	Protein fate
1372	type III effector HopAA1-1	Protein fate
1373	type III helper protein HrpW1	Protein fate
1374	type III chaperone ShcM	Cellular processes
1375	type III effector HopM1	Protein fate
1378	membrane-bound lytic murein transglycosylase D	Cellular processes
1379	type III transcriptional regulator HrpR	Regulatory functions
1380	type III transcriptional regulator HrpS	Regulatory functions
1381	type III helper protein HrpA1	Protein fate
1382	type III helper protein HrpZ1	Protein fate
1383	type III secretion protein HrpB	Protein fate
1384	type III secretion protein HrcJ	Protein fate
1385	type III secretion protein HrpD	Protein fate
1386	type III secretion protein HrpE	Protein fate
1387	type III secretion protein HrpF	Protein fate
1388	type III secretion protein HrpG	Protein fate
1389	outer-membrane type III secretion protein HrcC	Protein fate

1390	type III secretion protein HrpT	Protein fate
1391	negative regulator of hrp expression HrpV	Regulatory functions
1396	type III secretion protein HrcQb	Protein fate
1397	type III secretion protein HrcQa	Protein fate
1398	type III secretion protein HrpP	Protein fate
1399	type III secretion protein HrpO	Protein fate
1400	type III secretion cytoplasmic ATPase HrcN	Protein fate
1401	type III secretion protein HrpQ	Protein fate
1402	type III secretion protein HrcV	Protein fate
1403	type III secretion protein HrpJ	Protein fate
1404	RNA polymerase sigma factor HrpL	Transcription
1405	type III helper protein HrpK1	Protein fate
1406	type III effector HopB1	Protein fate
1408	protein of unknown function	
1409	conserved protein of unknown function	Cell envelope
1456	multicopper oxidase	Cellular processes
1514	protein of unknown function	
1568	type III effector HopAF1	Protein fate
1825	acetyl-CoA synthetase	Energy metabolism
1826	arginine/ornithine ABC transporter, periplasmic arginine/ornithine-binding protein	Transport and binding proteins
1832	acetylornithine delta-aminotransferase	Amino acid biosynthesis
1833	arginine N-succinyltransferase, alpha subunit	Energy metabolism
1844	carbon storage regulator	Regulatory functions
1886	L-serine dehydratase 1	Energy metabolism
1925	negative regulator of flagellin synthesis FlgM, putative	Regulatory functions
1949	flagellin	Cellular processes
1950	flagellin FlaG, putative	Cellular processes
1951	flagellar hook-associated protein FliD	Cellular processes
1987	CheW domain protein	Unknown function
2041	conserved protein of unknown function	Hypothetical proteins
2104	major facilitator family transporter	Transport and binding proteins
2105	thiamine biosynthesis lipoprotein,	Biosynthesis of cofactors, prosthetic

	putative	groups, and carriers
2130	DNA-binding response regulator, LuxR family	Signal transduction
2131	sensor histidine kinase	Signal transduction
2133	RNA polymerase sigma-70 family protein	Transcription
2136	2,4-diaminobutyrate 4-transaminase	Cellular processes
2137	MbtH-like protein	Unknown function
2150	pyoverdine sidechain peptide synthetase IV, D-Asp-L-Ser component	Cellular processes
2152	TonB-dependent siderophore receptor, putative	Transport and binding proteins
2160	efflux transporter, RND family, MFP subunit	Transport and binding proteins
2286	transcriptional regulator, GntR family	Regulatory functions
2287	methylisocitrate lyase	Energy metabolism
2288	2-methylcitrate synthase	Energy metabolism
2289	aconitase family protein	Energy metabolism
2367	ribose ABC transporter, periplasmic ribose-binding protein	Transport and binding proteins
2595	isochorismate synthase	Cellular processes
2597	yersiniabactin synthetase, salycilate ligase component	Cellular processes
2598	yersiniabactin synthetase, thioesterase component	Cellular processes
2599	yersiniabactin synthetase, thiazolinyl reductase component	Cellular processes
2600	yersiniabactin polyketide/non-ribosomal peptide synthetase	Cellular processes
2601	membrane protein, putative	Cell envelope
2602	yersiniabactin non-ribosomal peptide synthetase	Cellular processes
2603	ABC transporter, ATP-binding/permease protein	Transport and binding proteins
2604	ABC transporter, ATP-binding/permease protein	Transport and binding proteins
2605	TonB-dependent siderophore receptor, putative	Transport and binding proteins
2616	methyl-accepting chemotaxis protein	Cellular processes

2649	conserved protein of unknown function	Hypothetical proteins
2678	type III helper protein HopP1	Protein fate
2690	lipoprotein, putative	Cell envelope
2691	membrane protein, TerC family	Cell envelope
2696	mutT/nudix family protein	Unknown function
2701	fructokinase	Energy metabolism
2702	xylulokinase	Energy metabolism
2703	D-mannonate oxidoreductase	Energy metabolism
2705	mannitol ABC transporter, permease protein	Transport and binding proteins
2706	mannitol ABC transporter, permease protein	Transport and binding proteins
2707	mannitol ABC transporter, periplasmic mannitol-binding protein	Transport and binding proteins
2729	conserved protein of unknown function	Hypothetical proteins
2730	sulphite reductase	Central intermediary metabolism
2739	acyl-CoA dehydrogenase, putative	Fatty acid and phospholipid metabolism
2776	amino acid ABC transporter, permease protein	Transport and binding proteins
3087	type III effector HopAB2	Protein fate
3123	hypothetical protein	
3145	conserved protein of unknown function	Hypothetical proteins
3155	pyruvate ferredoxin/flavodoxin oxidoreductase family protein	Energy metabolism
3157	protein of unknown function	
3256	iron ABC transporter, periplasmic iron-binding protein, putative	Transport and binding proteins
3269	phosphate ABC transporter, periplasmic phosphate-binding protein	Transport and binding proteins
3291	methyl-accepting chemotaxis protein	Cellular processes
3364	isocitrate lyase	Energy metabolism
3455	3-oxoacid CoA-transferase, subunit A family	Energy metabolism
3466	alkanesulfonate monooxygenase	Central intermediary metabolism
3508	transcriptional regulator PsrA	Regulatory functions

3517	fatty oxidation complex, alpha subunit	Fatty acid and phospholipid metabolism
3544	conserved protein of unknown function	Hypothetical proteins
3582	catalase	Cellular processes
3598	dyp-type peroxidase family protein	Cellular processes
3648	acid phosphatase	Central intermediary metabolism
3699	methyl-accepting chemotaxis protein	Cellular processes
3882	polyamine ABC transporter, ATP-binding protein, putative	Transport and binding proteins
4001	type III effector protein AvrPto1	Protein fate
4024	ferredoxin--NADP reductase	Energy metabolism
4101	type III helper protein HopAK1	Protein fate
4155	argininosuccinate synthase	Amino acid biosynthesis
4168	glycerol kinase	Energy metabolism
4170	glycerol-3-phosphate dehydrogenase	Energy metabolism
4171	amino acid ABC transporter, periplasmic amino acid-binding protein	Transport and binding proteins
4172	amino acid ABC transporter, permease protein	Transport and binding proteins
4173	amino acid ABC transporter, permease protein	Transport and binding proteins
4248	3-hydroxyacyl-CoA-acyl carrier protein transferase	Fatty acid and phospholipid metabolism
4272	protein of unknown function	
4324	protein of unknown function	
4331	type III effector HopE1	Protein fate
4366	iron-regulated protein A, putative	Cell envelope
4367	lipoprotein, putative	Cell envelope
4381	conserved protein of unknown function	Hypothetical proteins
4555	conserved protein of unknown function	Hypothetical proteins
4560	outer membrane porin, OprD family	Transport and binding proteins
4588	type III effector HopS2	Cellular processes
4589	type III chaperone ShcS2	Cellular processes
4594	type III effector HopO1-2	Protein fate
4597	type III effector HopS1	Disrupted reading frame
4599	type III chaperone ShcS1	Cellular processes

4680	coronafacic acid synthetase, ligase component	Cellular processes
4681	coronafacic acid synthetase, acyl carrier protein component	Cellular processes
4682	coronafacic acid synthetase, dehydratase component	Cellular processes
4683	coronafacic acid beta-ketoacyl synthetase component	Cellular processes
4684	coronafacic acid synthetase component	Cellular processes
4685	coronafacic acid synthetase, ligase component	Cellular processes
4686	coronafacic acid polyketide synthase I	Cellular processes
4691	type III effector HopAD1	Protein fate
4703	type III effector HopAQ1	Protein fate
4704	DNA-binding response regulator CorR	Signal transduction
4705	sensor histidine kinase CorS	Signal transduction
4707	coronamic acid synthetase CmaD	Cellular processes
4708	coronamic acid synthetase CmaE	Cellular processes
4709	coronamic acid synthetase CmaA	Cellular processes
4710	coronamic acid synthetase CmaB	Cellular processes
4713	alanyl tRNA synthetase-related protein	Cellular processes
4718	type III effector HopAA1-2	Protein fate
4720	type III effector HopV1	Cellular processes
4721	type III chaperone ShcV	Cellular processes
4722	type III effector HopAO1	Protein fate
4723	conserved protein of unknown function	Hypothetical proteins
4727	type III effector HopG1	Protein fate
4733	protein of unknown function	
4776	type III effector HopI1	Protein fate
5016	proline permease	Transport and binding proteins
5017	bifunctional putA protein	Energy metabolism
5052	oxygen-independent coproporphyrinogen III oxidase, putative	Biosynthesis of cofactors, prosthetic groups, and carriers
5087	malonate decarboxylase, alpha subunit	Transport and binding proteins
5121	glutamate synthase, small subunit	Amino acid biosynthesis

5123	glutamate synthase, large subunit	Amino acid biosynthesis
5170	lipoprotein Blc	Cell envelope
5240	oxidoreductase, iron-sulfur-binding protein	Unknown function
5241	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	Biosynthesis of cofactors, prosthetic groups, and carriers
5255	carbonic anhydrase	Central intermediary metabolism
5279	cadmium-translocating P-type ATPase	Transport and binding proteins
5316	sulfonate ABC transporter, periplasmic sulfonate-binding protein, putative	Transport and binding proteins
5353	type III chaperone protein ShcA	Cellular processes
5354	type III effector HopA1	Protein fate
5391	outer membrane porin, OprD family	Transport and binding proteins
5393	conserved protein of unknown function	Hypothetical proteins
5432	conserved protein of unknown function	Hypothetical proteins
5465	acetyl-CoA hydrolase/transferase family protein	Unknown function
5472	DNA-binding protein HU family	DNA metabolism
5482	response regulator	Signal transduction
5499	aspartate ammonia-lyase	Energy metabolism
5510	oxaloacetate decarboxylase, alpha subunit	Transport and binding proteins
5511	acetyl-CoA carboxylase, biotin carboxylase, putative	Fatty acid and phospholipid metabolism
5517	conserved protein of unknown function	Hypothetical proteins
5542	periplasmic glucan biosynthesis protein, putative	Cell envelope
5622	PSPTO_5622	
5633	conserved protein of unknown function	

As expected, transcriptional activity differs considerably between cells cultured in MG and KB. Many genes that are induced in MG are related to the T3SS (Figure 3.9A, protein fate in Table 3.6). Genes involved in central intermediary metabolism, signal transduction and biosynthesis of cofactors are also up-regulated in MG. Transcriptional regulators show a mixed pattern, with some up-regulated (e.g., PSPTO_1404 *hrpL*) and others down-regulated (e.g., PSPTO_2133). Because expression of the T3SS is energy-costly, some of the changes may be related to a need to reduce the expression of housekeeping genes [225].

Unlike the ChIP-Seq experiment described earlier, RNA-Seq is not targeted to a specific sigma factor and thus it potentially provides information for all genes. Only nine percent of differentially expressed genes appear to be regulated directly by RpoN, based on the presence of RpoN binding and putative RpoN promoters upstream from them (Figure 3.9B). These genes are of particular interest because their expression occurs in concert with the induction of *hrp*-related genes in MG medium and they may have functions that are related to virulence and pathogenicity. The 12 genes in this class are reported in Table 3.7 and are differentially regulated with high confidence (p-value <= 0.001).

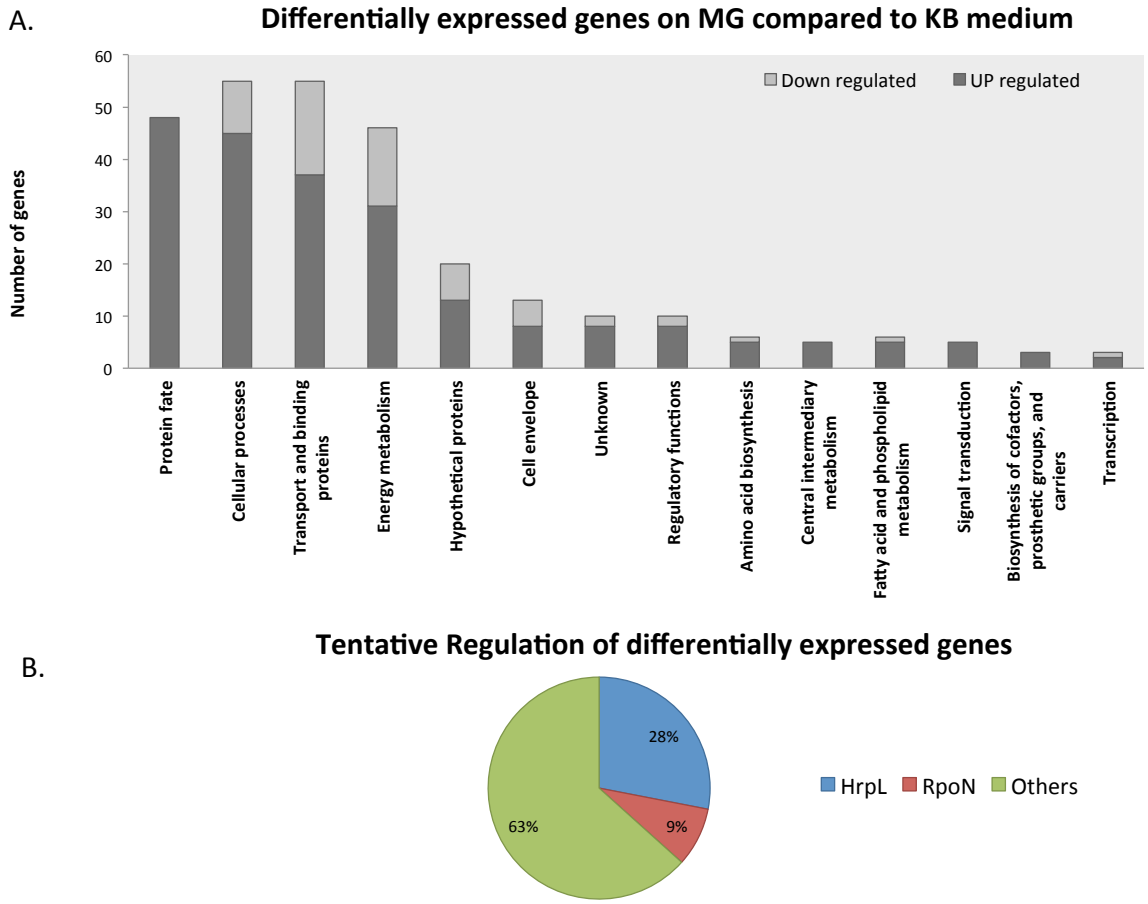


Figure 3. 9 Differentially expressed genes identified by RNA-Seq and their potential regulation

(A) The number of differentially expressed genes called by DE-Seq in categories. (B) Potential regulation of DE genes suggested by the presence of a promoter motif within 300 bases pairs upstream of corresponding genes.

Table 3. 7 List of differentially expressed genes that may be regulated by RpoN

"Op": gene in an operon;
"D": distance from 5' start site to RpoN motif for the first gene in an operon or distance from the second gene to the first gene for the second gene in an operon.
"Up": "up" indicates genes that are up-regulated in hrp-inducing conditions while "down" indicates genes that are down-regulated in hrp-inducing conditions compared to hrp-repressing conditions.

p-value	PSPTO	Potential function	Op	D		Up
0.00001	1255	amino acid ABC transporter, periplasmic amino acid-binding protein	1	12	TGGTACGGGCTTTGCT	up
	1276	glycine dehydrogenase	2	10	TGGCCGTAAAATTGCT	down
	1404	RNA polymerase sigma factor HrpL	1	12	TGGCATGGTTATCGCT	up
	4024	ferredoxin--NADP reductase	1	75	CGGAGAGTAAATCGCA	up
	4367	lipoprotein, putative	1	200	CGGCATTTTTTATGCC	down
	5511	acetyl-CoA carboxylase, biotin carboxylase, putative	1	20	GGGTGTCCCTGCGGCA	up
0.0001	1277	glycine cleavage system H protein	1	12	TGGCCGTAAAATTGCT	down
	3517	fatty oxidation complex, alpha subunit	1	24	TGGCTGATGGTGTGCA	up
	4155	argininosuccinate synthase	2	120	GGGCATGATGATTGCG	up
	5510	oxaloacetate decarboxylase, alpha subunit	2	102	GGGTGTCCCTGCGGCA	up
0.001	1826	arginine/ornithine ABC transporter, periplasmic arginine/ornithine-binding protein	1	-65/60	TGGCTTTTCAGGCGCT	down
	4171	amino acid ABC transporter, periplasmic amino acid-binding protein	1	12	TGGCACGACTCATGCC	up

5'-mRNA end capture

The RNA-Seq protocol was modified to enable the capture of 5'-mRNA ends as described in chapter 2. The modification is designed to specifically tag RNA molecules that are recovered from the cell with triphosphate residues at their 5'-ends, many of which are expected to be associated with transcription start sites (TSS) and promoters. Tagged reads were identified in the bulk RNA data using a simple string search and aligned separately to the DC3000 genome. The resulting profile has a distinctive "spiky" appearance (Figure 3.10). The 1000 captured ends supported by the most reads (referred to as the "top 1000" below) were analyzed further and the others were set aside.

Short genomic regions (40 bp) directly upstream from each captured 5'-end were extracted from the DC3000 sequence and analyzed using MEME [154] to search for conserved motifs. The recovered motifs are shown in Figure 3.11. Only one motif, a likely RpoD promoter (5'-TTG – N17-21 – TATAAT-3'), was obtained using data from cells grown in KB. In contrast, RpoD, HrpL (5'-GGAACC $-N_{16\text{-}17}$-CCAC-3'), RpoN (5'-tGG-N_9-tGC-3'), and FliA (5'- TcaaG-N_{14}-GccGANA-3') promoter motifs were obtained using data from cells grown in MG (Figure 3.11). The RpoN motif here should be compared with the one shown in Figure 3.5, which was derived using ChIP-Seq data. Some of the differences are probably attributable to the fact that the ChIP-Seq experiment sampled over 200 binding sites, whereas the 5'-end capture experiment sampled only 18 sites. The number of promoter motifs is less than the number reported by Filiatrault *et. al.* [62] in experiments where DC3000 was grown in MG (low iron) to late exponential phase (OD 0.6**;** over 10 hours). The difference may be due to the fact that the cells used here were harvested in an early phase

of bacterial growth, before nutrients and oxygen become limiting and sigma factors such as RpoE and RpoH become active.

Less than a quarter of RpoN promoters can be associated with TSSs in either MG or KB (Figure 3.12), where "associated" is defined loosely to mean that the TSS and promoter are located within 200bps from each other. A closer look at these regions reveals that only 13 are in close proximity to RpoN binding sites (Table 3.8). Among these, four appear to be induced by MG (PSPTO_4171, PSPTO_1255, *hrpL*, *shcV* differentially expressed by DE-Seq). Interestingly, genes related to ABC transporters are in this group. They may be good candidates for further study.

The distance between the RpoN motif and the closest captured 5'-end is well over 12 bp in many cases. These captured ends cannot be associated with the RpoN motif in any simple way as TSS. Whether RpoN plays a role in the transcriptional regulation for the neighboring genes is unknown. However, the vast majority of captured 5'-ends are not included in the top 1000 data set. A closer inspection of captured ends in these regions using the Artemis genome viewer shows captured 5'-ends with low read counts (above 10 read counts and higher than the surrounding background) downstream of RpoN motifs for nine genes (PSPTO_0110, *hopU1*, PSPTO_1015, PSPTO_1738, PSPTO_1921, PSPTO_1968, PSPTO_1977 (*flhF*), PSPTO_3386, and PSPTO_4296). It is possible that the cellular concentration of RpoN may be maintained at low levels [226] and that RpoN-mediated transcription is typically low except in unusual cases where high levels of expression are required. Survival is reduced when RpoN is over-expressed [227]. In fact, in contrast to HrpL-FLAG, we were unable to detect RpoN-FLAG protein on a Western Blot in spite of much effort (data not shown) even though its activity was detectable (Figure 3.3).

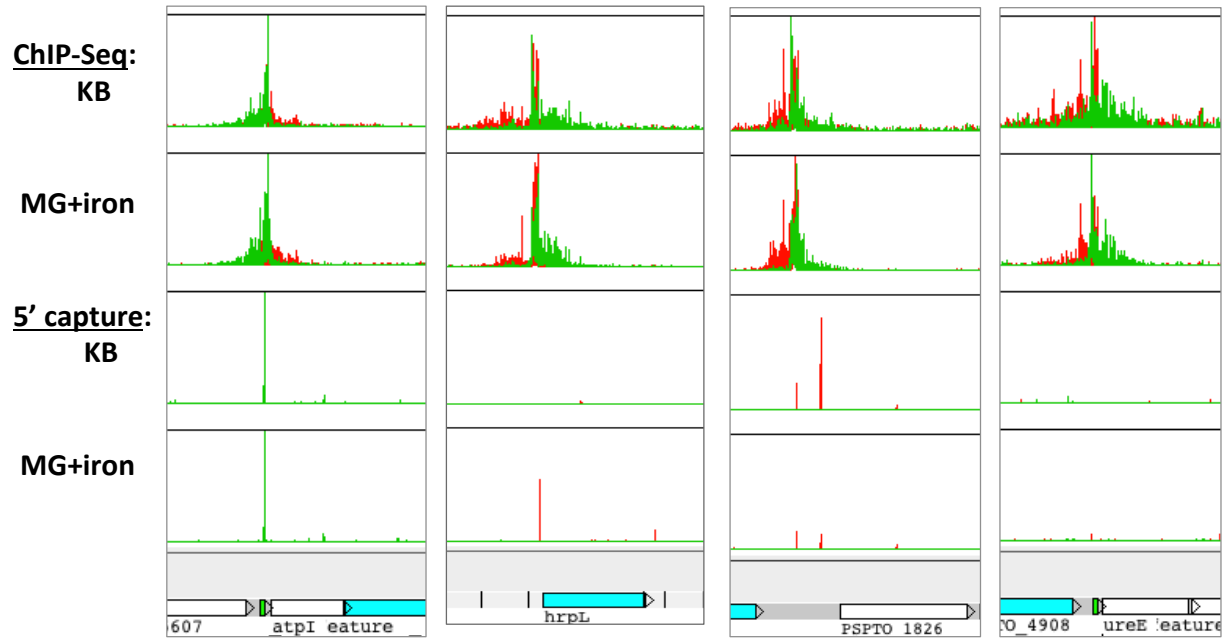


Figure 3. 10 Example of ChIP-Seq peaks and captured ends from 5' capture data.

Four chromosomal regions are shown, aligned with the corresponding ChIP-Seq and 5'-end capture results for two culture conditions. Red and green traces correspond to sequence reads mapping to the positive or negative strands, respectively. As noted in the text, the ChIP-Seq results are virtually identical in both conditions. In contrast, transcriptional activity as gauged by captured ends, varies with the culture condition. Note that most genome coordinates have zero reads associated with them (99.61% and 99.49% on the positive strand on KB and MG respectively; 99.50% and 99.36% on the negative strand on MG and MG respectively).

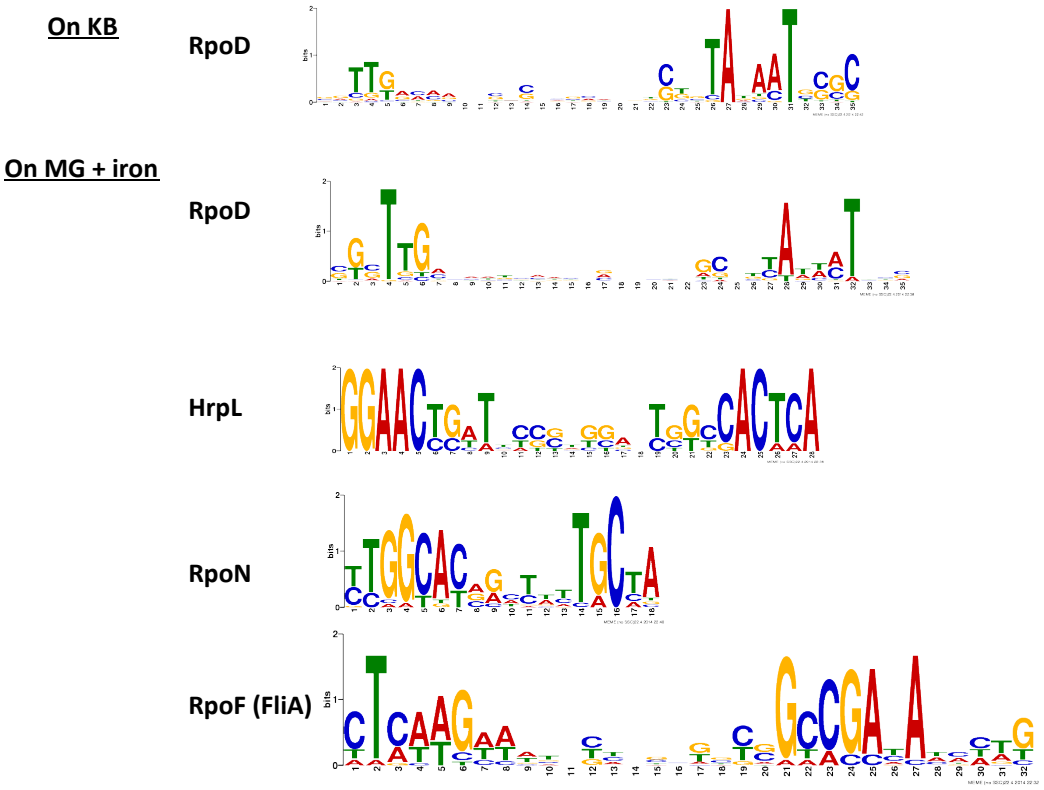


Figure 3. 11 Motifs identified by MEME from 5' capture data

Motif identified by MEME using 5' capture data from KB (A) and MG (B)

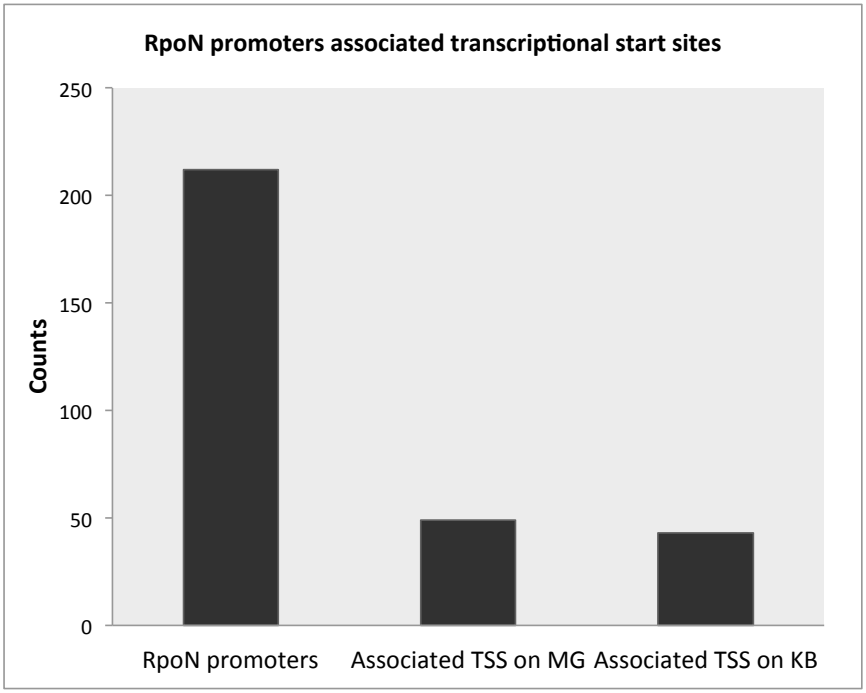


Figure 3. 12 Distribution of RpoN promoters with respect to nearby transcriptional start sites

Table 3. 8 List of sequences of RpoN motifs called by MEME in 5' capture data

"RpoN motifs called by MEME": all motifs called by MEME that participated in the RpoN sequence logo shown in Figure 3.11
"genes": genes immediately downstream of RpoN motif
"Function": annotated function of the downstream genes
"RpoN binding site": "yes" indicates that there is RpoN binding site (inferred from ChIP-Seq data), suggesting the RpoN motif is authentic; "no" indicates no RpoN binding site in vicinity of called RpoN motif, suggesting a mistakenly called motif
"Number of TSS": the number of TSS in the vicinity of the RpoN motif.

No	RpoN motifs called by MEME	Genes	function	RpoN binding site	number of TSSs
1	TTGGCACAGCTTCTGCTA	5668	small RNA	yes	1
2	CTGGCACCGTTGTTGCTA	1957	fliE	yes	1
3	TCGGCACAAATTATGCCA	5606	aptI	yes	1
4	*TTAGTACAGATATTGCTA*	*4721*	*shcV*	*no*	1
5	CTGGCATGGTTATCGCTA	1404	hrpL	yes	1
6	CCGGCATCGCTATTACAA	5439	Rhs family protein, truncation	yes	3
7	TTGGCACGACCTATGCTT	927	type IV pilus biogenesis protein	yes	1
8	*TTCGCACCACCACTGCCA*	*4805*	*PhoH-like protein*	*no*	1
9	CTGGCGCAAACCTTGCTA	3129	conserved protein of unknown function	yes	1
10	TTGGCACGACTCATGCCC	4171	amino acid ABC transporter, periplasmic amino acid-binding protein	yes	1
11	TTGGCATAGAGGGTGCTA	4108	high-affinity branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein	yes	1
12	CTGGTTCGGAAATTGCTA	4136	mino acid ABC transporter, periplasmic amino acid-binding protein	yes	1
13	*TTGACACGCTCTCTGCCA*	*5147*	*polyhydroxyalkanoate granule-associated protein PhaF*	*no*	1
14	GTGGCACACTTTCTGCTG	1255	amino acid ABC transporter, periplasmic amino acid-binding protein	yes	1
15	TTGGCACAGCTTCTGCTA	5669	small RNA	yes	11
16	GTGGCACACTTTCTGCTG	2775	amino acid ABC transporter, periplasmic amino acid-binding protein	yes	1

Returning to the genes of interest in Table 3.7, only a quarter of RpoN promoters (described above) are associated with captured 5'-ends at a distance appropriate for TSS. Overall, PSPTO_1255 and PSPTO_4171 (amino acid ABC transporter, periplasmic amino acid-binding protein) are the most interesting candidates for virulence factors because they are differentially expressed (up-regulated in *hrp*-inducing conditions), the distance between their TSSs and RpoN motifs are plausible for active promoters, and their functions perhaps suggest a role in transporting nutrients to bacterial cells.

RpoN regulates motility in DC3000

In *P. aeruginosa*, the genes involved in flagella synthesis have been divided into four classes, suggested by the work of Dasgupta *et. al*. [228]. Class I genes encode the regulators FleQ, a transcriptional regulator containing a σ^{54} interaction domain [229], and FliA, an alternative sigma factor required for flagellin synthesis [230]. The expression of class I genes are primarily regulated by factors outside of the flagella regulon [228]. Class II genes encode structural components of the basal body, P ring, MS ring, filament cap protein and export apparatus. They depend on RpoN and FleQ for transcription. Class III genes require FleR and RpoN for transcription and provide components for the basal body rod, L ring, and hook-filament junction. FleR encodes a σ^{54} dependent transcriptional regulator with a response regulator receiver domain. FleS encodes a sensor histidine kinase. Upon sensing environmental or cellular signals, FleS transduces the signal to FleR and FleR becomes active. Together with RpoN, FLeR/S regulates the class III genes. Finally, class IV genes are dependent on the FliA sigma factor, and encode filament proteins and control filament length.

As discussed above, RpoN is involved in flagella biosynthesis in many bacteria [113, 115, 177-183, 185]. In DC3000, our ChIP-Seq results and motif search suggest that RpoN regulates Class II and III flagella genes (Figure 3.13). This is consistent with the model proposed in *P. aeruginosa* [228]. Class II flagella genes include *flhF*, *fliE* (*fliE*, *fliF*, *fliG*, PSPTO_1960, *fliI*, and PSPTO_1962 are in the same operon), *fliL* (*fliL*, PSPTO_1968, and *fliM* are in the same operon), *flhA*, *fleS* (*fleS*, and *fleR* are in the same operon). Class III flagella genes include *flgB* (*flgB*, *flgC*, *flgD*, *flgE-1*), *flgF*, and *fliK*. RpoN may also regulate other genes in flagella gene clusters including PSPTO_1986 (*parA*) and PSPTO_1926 (flagella protein) (Table 3.4).

Besides well-known components of flagella regulated by RpoN, there might be other factors involved in motility to be found. For example, a small RNA, located in the same operon with flagella genes, was shown to affect swimming in *Vibrio cholera* [115]. The dependence on RpoN of genes in flagella biosynthesis, and/or type IV pilus (PSPTO_0927) may account for the reduction of swarming activity observed in our data (Figure 3.3C). Swarming dependent on RpoN was reported for *P. aeruginosa* [231, 232].

Besides RpoN, other sigma factors can be involved in regulation of motility in DC3000. σ^{70} promoter motifs have been identified upstream of *fleQ* and *fliD* (flagellar hook-associated protein) [62]. The FliA promoter motif is found upstream of *fliC*, *motA-2*, *flgK* and *flgM* [62]. This is not unexpected as motility is involved in many aspects of bacterial lifestyle and thus need more complex regulation [233, 234].

RpoN regulates small noncoding RNAs in DC3000

The role of small noncoding RNAs (sRNAs) in bacterial gene regulation has attracted increasing attention. The Diversity and function of sRNAs in bacteria has been summarized in recent reviews [235-238]. sRNAs can be cis-encoded, trans-encoded or encoded within the 5'-untranslated regions of protein coding regions [235]. There are three major classes of sRNAs involved in the adaptive response to environmental stress and stimuli in *P. aeruginosa* [238]. The first class acts by base-paring with mRNAs to control the stability and/or translational output of the target mRNAs. This class usually requires the presense of Hfq, exemplified by *RsmY/Z*, which targets the RsmA mRNA to control processes in chronic and acute infection [239, 240]. The second class of sRNAs binds to proteins and modulates their activity. An example is the binding of *crcZ* to Crc protein in order to regulate the response to a poor carbon source such as mannitol [241]. Lastly, some sRNAs are encoded by clustered regularly interspaced short palindromic repeats (CRISPR). These RNAs can target double or single-stranded DNA [242]. We will focus here mainly on RpoN regulated small RNAs.

In *P. aeruginosa* PAO1, the nitrogen-regulated small RNA *NrsZ* positively regulates *rhlA,* a gene involved in synthesis of rhamnolipid surfactants, at the post-transcriptional level. Consequently, *NrsZ* modulates swarming motility of the bacteria. It was also shown that *NrsZ* transcription is controlled by RpoN and the two-component system NtrB/C [243].

In *Vibrio cholera* O37 strain V52, a small RNA, *flaX*, is involved in regulation of motility [115]. This small RNA is in the same operon with the flagellin *flaA* gene. The phenotype was complemented by *flaX* expression in *trans* suggesting a genuine regulation of *flaX* in motility of the *Vibrio cholera*.

The small RNAs *GlmY* and *GlmZ* of *Yersinia pseudotuberculosis* are regulated by RpoN and the two-component system GlrR/GlrK as well as IHF [244]. *GlmY* acts upstream of *GlmZ* to activate expression of the *glmS* gene, which encodes glucosamine-6-phosphate (GlcN6P) synthase GlmS [245]. GlmS catalyzes formation of GlcN6P, which initiates the pathway that generates peptidoglycan and lipopolysaccharide, precursors of cell wall synthesis [246].

In DC3000, *crcX* and *crcZ* play a role in carbon source utilization [217, 247] and most likely depend on RpoN, the CbrA/CbrB two-component sensor-regulator system, and perhaps other regulators for transcription [217]. These sRNAs may bind to Crc proteins and consequently sequester Crc by antagonism [247]. Free Crc proteins are believed to repress the translation of carbon metabolism genes by binding to their mRNAs, probably with the involvement of Hfq.

Our ChIP-Seq and RNA-Seq data suggest that at least three closely related small noncoding RNAs (*crcX*, *crcY*, and *crcZ*) are regulated by RpoN, consistent with the results reported by Filiatrault et al. [217]. Two of the ncRNAs appear to be functional (PSPTO_5668, namely *psr1* or *crcZ*, and PSPTO_5669, namely *psr2* or *crcX*) while one is disrupted by a transposon insertion in DC3000 (*psr3*, also namely snr0106 or *crcY*). In addition, two other RpoN binding sites are in regions that potentially could encode a small RNA. One is adjacent to PSPTO_4108, encoding a periplasmic high-affinity branched-chain amino acid ABC transporter. The other is located in the intergenic space between PSPTO_0287 and PSPTO_0288 (Figure 3.14). Whether these regions encode small RNAs remains to be demonstrated.

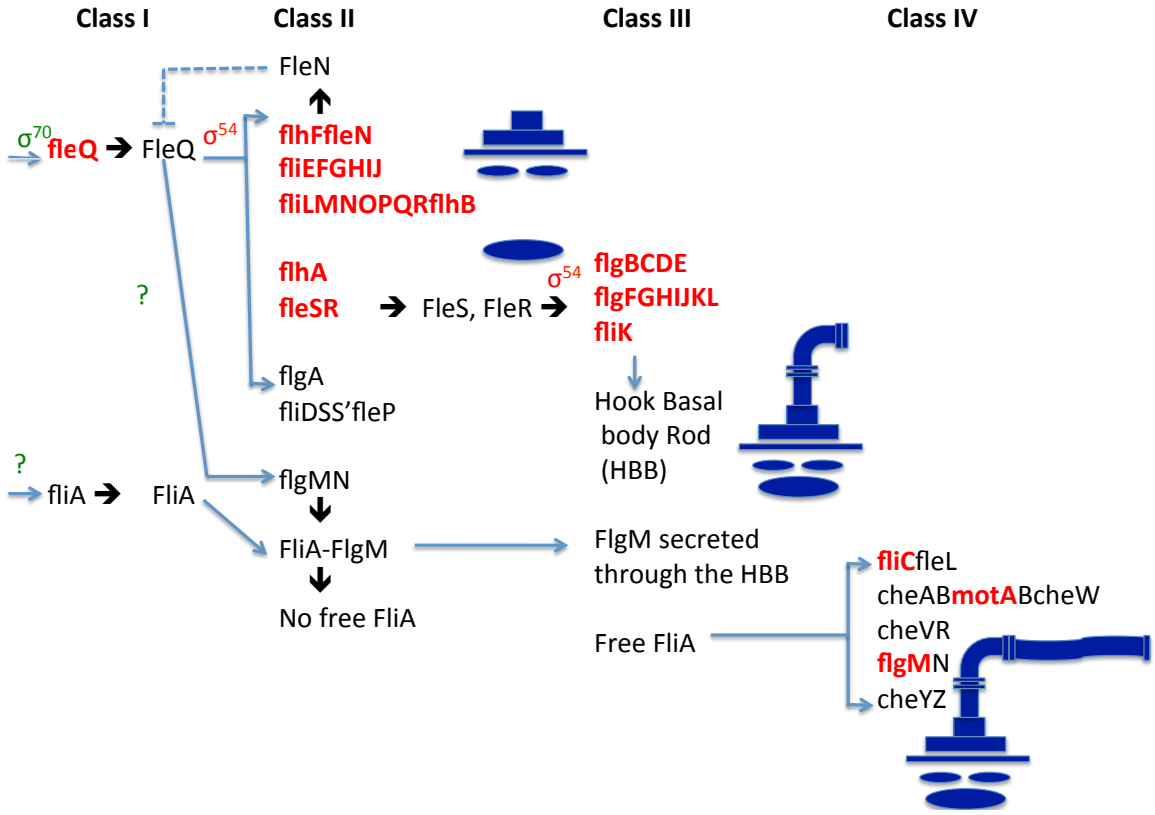
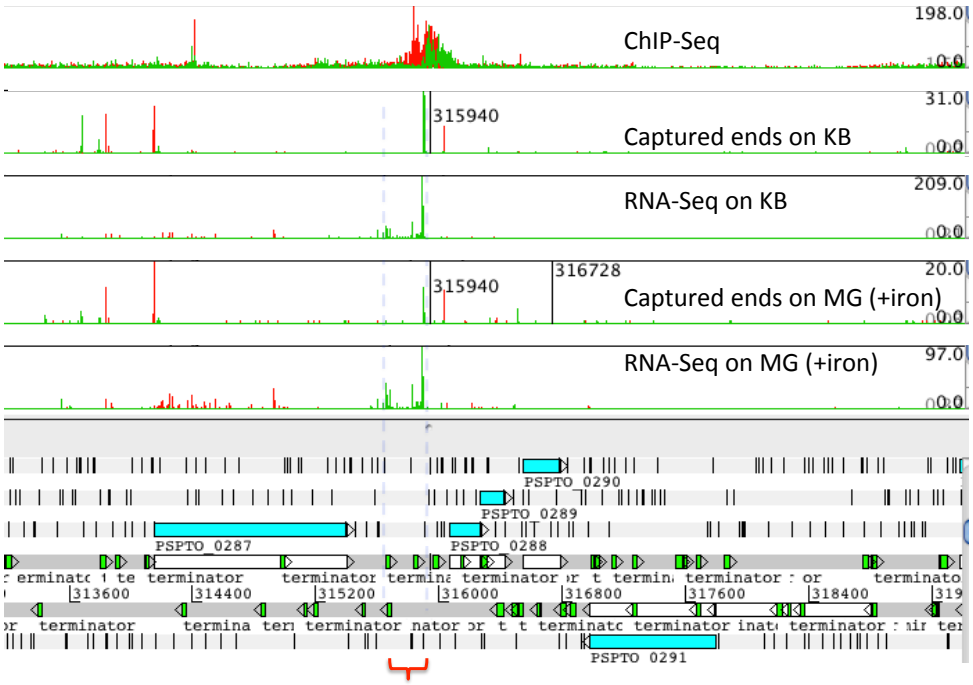


Figure 3. 13 RpoN coregulates flagella biosynthesis

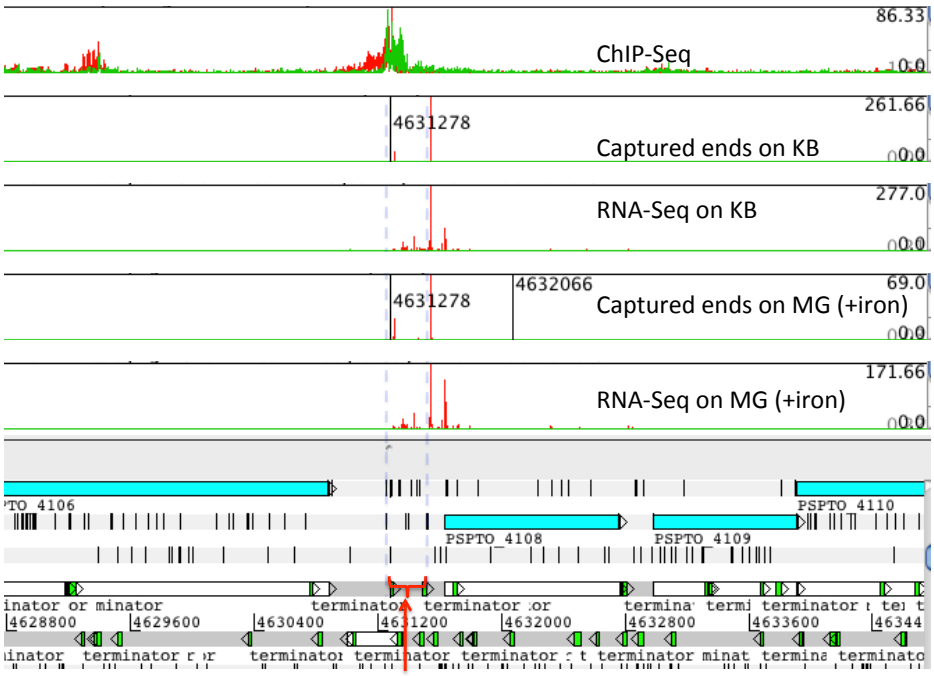
Genes in flagella biosynthesis have been classified into 4 groups [228]. Genes with evidence of regulation by the indicated sigma factors identified in ChIP-Seq and/or promoter search are marked in red.

A.



Potential sncRNA (315561..315910)

B.



Potential sncRNA (4631309..4631506)

Figure 3. 14 Potential small non-coding RNAs regulated by RpoN

Rho-independent transcriptional predicted by TransTermHP [248] were overlaid onto the DNA sequence to help identify possible transcript margins. (A) Potential sncRNA at coordinate (315561..315910). (B) Potential sncRNA at coordinate(4631309..4631506).

Enhancer binding site upstream of *hrpL*

The role of EBPs in activation of σ^{54}-dependent transcription has been reviewed in detail by Bush and Dixon [107]. Enhancer proteins often function as hexamers that are formed upon sensing environmental signals or which exist in a preformed state. EBPs bind to upstream activator sequences (UAS), usually 80 to 150 base pairs from the transcription start site. Although the one-dimensional distance between the enhancer binding site and promoter is fairly large, the EBPs are brought into proximity with the σ^{54} binding site via DNA looping by other factors such as integration host factor (IHF). When ATP is supplied, EBPs undergo conformational changes leading to remodeling of σ^{54} which facilitates the transition from the holoenzyme-promoter closed complex to an open complex to initiate transcription [249].

HrpL requires RpoN and the HrpR/S EBPs for transcription [194, 195]. The DNA region that binds HrpR/S is located –147 to –101 bp upstream from TSS position +1 [75]. As noted earlier, our ChIP-Seq results clearly demonstrate RpoN binding in the vicinity of the RpoN promoter that controls transcription of *hrpL*. In addition, the mRNA 5'-end capture results are consistent with transcription from this promoter at the +1 position, 12 bases downstream from the RpoN motif in cells grown under *hrp*-inducing conditions. In contrast, the 5'-capture signal is absent in cells grown in rich medium, which suppresses expression of the T3SS.

Since EBPs interact closely with the polymerase holoenzyme to create an open complex prior to transcription [107], in principle the ChIP-Seq experiment might detect RpoN binding to the promoters as well as to the EBPs binding to sequences further upstream at the UAS. However, we saw little evidence of an unambiguous enhancer signal

in our data at the *hrpL* locus. To examine enhancer binding at this location using a different method, a plasmid-based assay was devised in which the *hrpL* promoter, along with varying lengths of genomic sequence adjacent to it, was cloned upstream of a GFP reporter. The results show that a region extending from –166 to –111 bp from the +1 position are required for GFP expression (Figure 3.15 A-B). When combined with the results from [75], the HrpR/S enhancer appears to bind in the region from -147 to -111. The ChIP-Seq profile at the *hrpL* locus shows a small perturbation in this region (Figure 3.15 C) upstream from the main peak (attributable to RpoN binding). If this subtle feature is due to enhancer binding, it suggests that it might be possible to refine the ChIP-exo protocol to reveal them more clearly. Interestingly, a motif found upstream of genes expressed in a HrpR/S dependent manner (5'-GA-N-GTTTT-3'; [225]) is present in this region at coordinates 1542668..1542675 (GATGTTTT).

There are 22 candidate EBPs in DC3000, based on the presence of σ^{54} interaction domains and the characteristic amino acid motif "GAFTGA" essential for interaction between EBPs and the sigma factor [215]. Half of them appear to possess a regulatory domain capable of receiving signals, but the others do not (Table 3.9). RpoN may regulate the transcription of two EBPs, PSPTO_1956 (*fleR*) and PSPTO_0352 (nitrogen regulation protein NR(I)) based on the ChIP-Seq data and promoter search (Table 3.9). Both have regulatory domains. These EBPs might be regulated at transcriptional level. For EBPs without regulatory domains such as HrpR/S, environmental signals may influence their transcription, or they may interact with other factors that sense environmental signals directly.

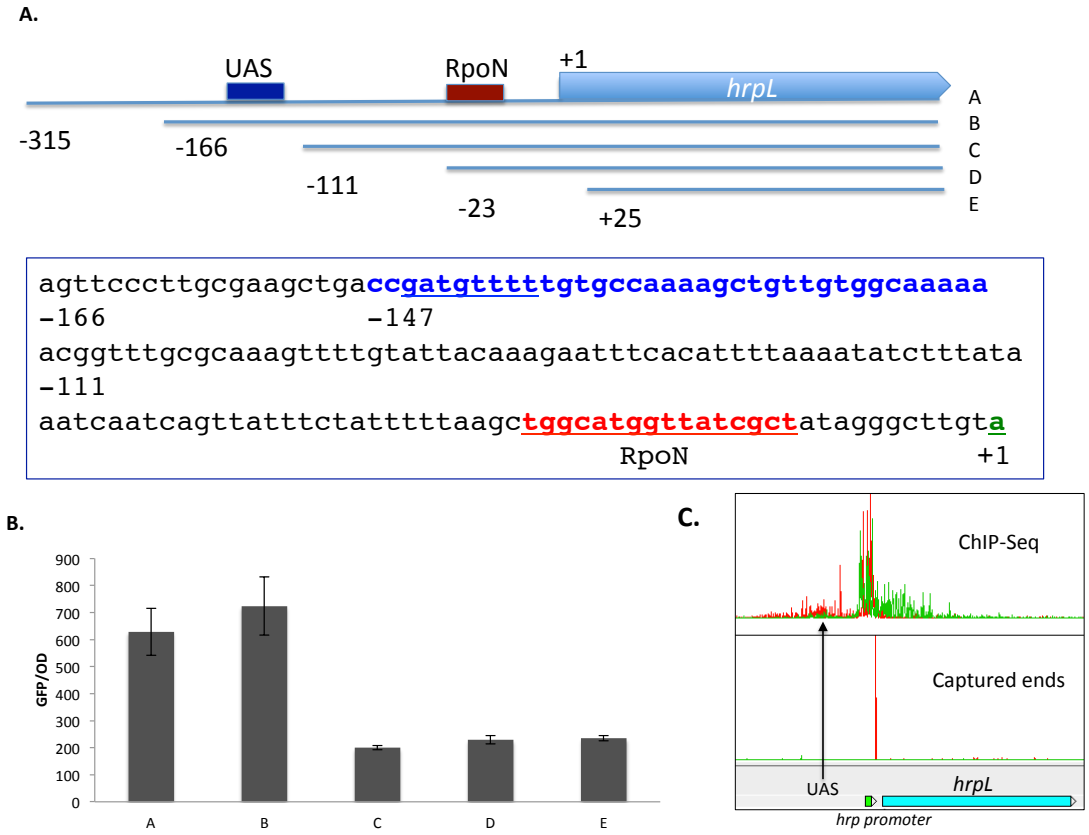


Figure 3. 15 *hrpL* promoter region

(A) Different lengths of DNAs upstream from *hrpL* containing potential regulatory sites were cloned into reporter constructs. The actual sequence is shown for reference. The HrpR/S binding site is marked in blue, the RpoN motif is in red and the TSS is in green (+1). (B) Relative GFP/OD for 5 reporter constructs. High expression is observed for fragments containing potential regulatory elements (constructs A and B) but low expression for constructs lacking one or more elements (constructs C, D and E). Values are the average of 3 replicates. (C) ChIP-seq and 5' capture data for the region surrounding *hrpL*. Sequence reads mapped to the positive strand of chromosome are in red while those mapped to the negative strand are in green.

Table 3. 9 List of EBPs in DC3000.

This list is modified from Francke *et al*. [215]. Notes on σ^{54} regulation are based on the presence of RpoN motifs and/or RpoN binding sites (identified by promoter search or ChIP-Seq). '+': RpoN motif is present upstream of the gene. '?': RpoN binding site is present upstream of the gene but the motif oriented in the opposite direction from transcription.

PSPTO	Function	σ^{54}	Protein domains		
			Signaling/signal sensor	σ^{54} binding	DNA-binding
5217	σ^{54} binding protein			PF00158	
1380	type III transcriptional regulator HrpS			PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
1379	type III transcriptional regulator HrpR			PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
3467	σ^{54} dependent transcriptional regulator, putative			PF00158	
1280	transcriptional regulator TyrR, putative		PAS 8	PF00158	
5424	σ^{54} dependent transcriptional regulator			PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
2549	σ^{54} dependent transcriptional regulator			PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
1823	phenylalanine hydroxylase transcriptional activator PhhR			PF00158	
3741	σ^{54} dependent transcriptional regulator	?		PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
3046	σ^{54} dependent transcriptional regulator	?		PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
2259	σ^{54} dependent transcriptional regulator			PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
111	σ^{54} dependent transcriptional regulator	?	PF08448, PAS 4, PASfold	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
1954	transcriptional regulator FleQ		PF06490, Flagellar regulatory protein FleQ	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
2951	σ^{54} dependent transcriptional regulator	?	PF06505, XylR N, Activator ofaromatic catabolism; PF02830, V4R domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
4176	σ^{54} dependent transcriptional regulator/response		PF00072, Response regulator receiver	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family

	regulator		domain		
4292	σ^{54} dependent transcriptional regulator/response regulator		PF00072, Response regulator receiver domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
823	type 4 fimbriae expression regulatory protein pilR		PF00072, Response regulator receiver domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
334	alginate biosynthesis transcriptional regulatory protein AlgB		PF00072, Response regulator receiver domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
5399	σ^{54} dependent transcriptional regulator/response regulator		PF00072, Response regulator receiver domain	PF00158	
1956	σ^{54} dependent transcriptional regulator/response regulator FleR	+	PF00072, Response regulator receiver domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family
964	σ^{54} dependent transcriptional regulator/response regulator		PF00072, Response regulator receiver domain	PF00158	
352	nitrogen regulation protein NR(I)	+	PF00072, Response regulator receiver domain	PF00158	PF02954, HTH_8, Bacterial regulatory protein, Fis family

RpoN may negatively auto-regulate

A prominent ChIP-Seq signal is observed immediately upstream from the RpoN gene itself (Figure 3.16). This region contains two divergent RpoN promoter motifs, with one potentially positioned to support transcription of RpoN (70 base pairs upstream from the *rpoN* predicted translational start site) and the other positioned to support antisense transcription for PSPTO_4452. It is not clear that either of the two RpoN motifs is a functional promoter. However, results in other bacteria suggest that RpoN may in some way regulate its own transcription. In *P. putida*, expression of RpoN is tightly controlled [250]. Multiple copies of RpoN lower bacterial survival of *Brucella abortus* and *Rhizobium leguminosarum* [251, 252]. Negative auto-regulation of RpoN was also reported in *Geobacter sulfurreducens* [68]. A strain in which RpoN was over-expressed showed reduced binding activity at RpoN targets by twofold in ChIP-qPCR in *P. putida* and *G. sulfurreducens* [109, 250].

One characteristic of negative auto-regulation is that it shortens the time between the reception of a signal and the transcriptional response [253]. It has been shown that negative auto-regulation reduces the rise-time (the time to reach half steady-state level), but does not affect the turn-off time (the time to reach half steady state level after transcription turns off) [253]. This system may be helpful in governing RpoN function in bacteria. As a central regulator, RpoN controls many processes including T3SS. It may be necessary to quickly increase RpoN levels to cope with environmental changes, while avoiding overexpression that will imbalance the cell's physiology.

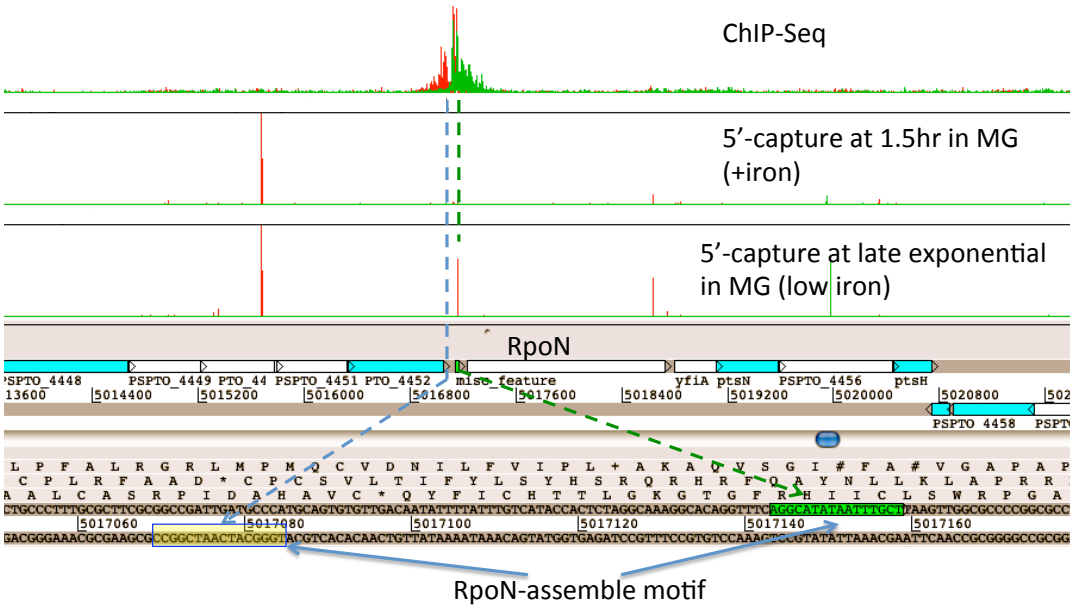


Figure 3. 16 RpoN motifs upstream of *rpoN* gene

Profiles of ChIP-Seq, 5'capture in this experiment, and 5' capture in [62] are shown above gene annotation and promoter motifs. Sequence reads mapped to the positive strand of chromosome are in red while those mapped to the negative strand are in green.

Conclusion

The experiments described here provide a preliminary assessment of the RpoN regulon in DC3000 and its activity under two experimental conditions. In addition to confirming its well-known role in the regulation of *hrpL*, the results suggest that RpoN regulates flagella biosynthesis, energy metabolism, nitrogen metabolism, transport and binding proteins, regulatory factors including enhancer binding proteins, and the transcription of small RNAs (Table 3.4). Active expression was observed for small RNAs *crcX* and *crcZ* (involved in carbon source utilization), the T3SS (as expected via HrpL) and the flagella system.

However, other regulon members appeared to be unexpressed under our experimental conditions even though the RpoN binding profile remains virtually constant between conditions, exemplified by the RpoN binding site upstream of *ureE* (PSPTO_4909). The results are consistent with the current model for RpoN-mediated transcription in which the sigma factor (along with RNA polymerase) is thought to form a stable complex with DNA in the absence of enhancer binding proteins [198]. If the model is correct, then the ChIP-Seq experiment has probably probed the vast majority of RpoN promoters even though many of them are transcriptionally inactive. Activation of these promoters will require identifying the appropriate growth conditions or manipulating the EBPs to force activation (see [108]).

The extent to which RpoN regulates genes involved in virulence and pathogenicity remains unknown. The T3SS is a well-known virulence determinant regulated by RpoN. It was hypothesized that RpoN may regulate both coronatine biosynthesis and T3SS [195, 254, 255]. The dependence of coronatine synthesis on RpoN may be indirect via the

dependence of *corR* and *corS* on a *hrp* promoter upstream of the operon of hopAQ1 [256]. RpoN may control Type IV pili (PSPTO_0927), which are required for virulence in *P. syringae*. pv. *tabaci* [176] and *acidovoax avenae*, a fruit blotch pathogen of cucurbits [257]. RpoN may regulate *hcp-1* and *hcp-2,* which encode components of the Type VI secretion system [110, 189]. The T6SS was shown to be involved in host infection in *P. aeruginosa* [110], but whether *hcp-1* and *hcp-2* have any role in DC3000 virulence remains to be tested.

Acknowledgement

We thank HoangChuong BuiNguyen for assistance in tagging *rpoN* with the FLAG epitope, Christopher R. Myers for matching DC3000 genes with role categories from the JCVI website, Paul V. Stodghill for comments on sequence analysis procedures, Bryan Swingle for comments on EBPs, and Melanie J. Filiatrault for comments on sncRNAs.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

Introduction

The major conclusions of the HrpL and RpoN regulon projects have been described in the preceding chapters. In this section, we will explore data that we found to be particularly interesting and will also propose experiments that might be conducted in the future.

Regulation of the type III secretion system

In the course of our research, 73 HrpL binding sites associated with *hrp* promoter motifs were identified in a ChIP-Seq analysis. This list includes 52 of the 54 annotated *hrp* promoters and 21 new candidates. In most cases, there are plausible transcription start sites adjacent to the *hrp* promoters and the genes downstream from them are expressed preferentially under *hrp*-inducing conditions. This, along with other experimental evidence described in detail in chapter 2, strongly supports the addition of the 21 new promoters to the annotated set, significantly enlarging the boundaries of the HprL regulon for both promoters and the genes directly under their control.

HrpL regulon

Experimental and computational searches for *hrp* promoters, virulence genes and host related genes have been carried out intensively for many years [48, 65, 80, 82, 83, 85,

86, 130, 173, 258]. New candidate virulence genes are found infrequently, suggesting that the list of virulence factors is reaching saturation. However, it is important to note that genetic redundancy has made it a challenge to identify new genes by mutational analysis. This problem has been particularly important in the search for effectors [166]. More recently, the combinatorial deletion of effectors has been used to reveal groups of effectors whose functions are collectively essential but individually dispensable [89]. The minimal functional repertoire approach, which aims at removing all functional effectors and reassembling them in various combinations to explore the composite roles of T3SS effectors, has also become an important strategy [90].

The expanded HrpL regulon described in chapter 2 includes only one new effector, PSPTO_5633 (*hopBM1*). This gene will be deleted in the next generation of "effectorless" DC3000 strains (Alan Collmer, personal communication). Nothing is known yet about how PSPTO_5633 functions *in planta*, and it joins a list of other effectors whose functions are not understood. However, this protein may have another property that is equally interesting. Although PSPTO_5633 behaves as a typical effector in the standard Cya-based, T3SS-dependent translocation assay, its N-terminal peptide sequence is unusual because it resembles the signal sequence recognized by the Sec secretion system. Is it possible that this protein is also exported from the cell by the Sec pathway? Our experiments did not eliminate this possibility because they were designed to determine whether a *sec* mutant blocked translocation into the plant cell (an unlikely scenario) rather than whether the mutant blocked secretion itself. Therefore, there is a possibility that PSPTO_5633 might be exported via T2SS into the medium (or into the plant apoplast). Because a small number of

other T3SS-related proteins have similar N-terminal peptide sequences (Table 4.1), this question is worth a closer look.

Our data also revealed that HrpL regulates a number of genes that are not obviously related to the Type III secretion system. Two genes in this group were reported previously. PSPTO_0834, an alcohol dehydrogenase, strongly influences bacterial virulence [88] while PSPTO_2130, a potential DNA-binding response regulator, did not show a virulence phenotype when the deletion mutant on DC3000 was tested ([130] and also in our study). In addition, the deletion of six other new regulon members individually from DC3000 revealed no detectable virulence phenotypes. However, one of these genes, the protease inhibitor *aprI* (PSPTO_3331), has been reported to increase plant resistance to DC3000 when it is expressed *in planta* [171]. The role of AprI in the bacterial cell is not understood. AprI might help bacteria protect their flagellin from the activity of the AprA protease, which is thought to degrade flagellin in order to evade the plant immune response. However, AprI is thought to be localized in the periplasmic space and presumably does not contact flagellin outside of the cell.

The lack of a connection to the T3SS for many new regulon members, along with the lack of virulence phenotypes when they are deleted, raises the question as to why they are in the regulon at all. This question is difficult to answer but two observations may be helpful. First, many (but not all) new regulon members are present in other *P. syringae* strains accompanied by *hrp* promoter motifs. The previously known regulon members exhibit the same general pattern. This suggests that the inclusion of these new members in the regulon—resulting in their coordinated expression with the T3SS—offers some advantage related to bacterial growth in the plant. Second, a subset of the new regulon

members is highly conserved among the *Pseudomonads*, even in non-pathogenic strains. This subset may therefore be involved in functions unrelated to virulence. It is possible that the HrpL regulon supports plant-microbe interactions in a more general way than previously suspected.

Finally, it is not absolutely certain that all HrpL regulon members have been identified. Although the high-throughput methods we used are very sensitive, arbitrary thresholds were imposed to select working subsets for analysis from the ChIP-Seq and RNA-Seq data. It is possible that *bona fide* HrpL regulon members were excluded. In fact, by lowering the thresholds it is possible to obtain additional candidates. Therefore, in the future, it may be advantageous to revisit the data sets, especially if new evidence suggests that a specific gene should be investigated as a possible regulon member. On the other hand, a blind approach (in which thresholds are lowered to bring in large numbers of new candidates) is likely to recover increasing numbers of false positives rather than quality candidates, and would probably be very inefficient.

Table 4. 1 N terminal signal peptide analysis for selected T3SS-related proteins

Prediction is based on Effective database [259]:

[a] "+" if SignalP detected a signal peptide for Sec-pathway secretion.

[b] result from Neural Network Predictor, a score is between [0 ... 1.0]. The higher the score, the more confident is the prediction.

[c] result from Hidden Markov Model Predictor, a score is between [0 ... 1.0]. The higher the score, the more confident is the prediction.

[d] "+" if the predicted result supports a Type III secreted protein.

GREEN background: the positive prediction for the corresponding secreted pathway.

Protein	**Is sec secreted**			**Is T3 secreted**		***hrp* promoter**
	+/- [a]	**NN[b]**	**HMM[c]**	**+/- [d]**	**Effective T3 score**	
PSPTO_5633	+	0.753	1		0	This study
PSPTO_3331	+	0.643	0.998		0	This study
PSPTO_3332		0	0	+	1	
HrcJ	+	0.284	0.97		0	Annotated, in operon
HrcC	+	0.808	0.998		0	Annotated, in operon
HrpT	+	0.582	0.999		0	Annotated, in operon
HopAJ1	+	0.78	1		0	Annotated
PSPTO_0524	+	0.432	1		0	Annotated
PSPTO_5622	+	0.899	1		0	Annotated

RpoN regulon

RpoN (σ^{54}) is an unusual global regulator in prokaryotes. The requirement for enhancer binding proteins for transcriptional activation effectively divides the regulon into multiple sub-regulons whose activation depends on various signals. The most recent model for RpoN promoter interaction suggests that RpoN holoenzyme binds to promoters even when EBPs are unavailable and can remain bound without initiating transcription. Our results are consistent with this model. First, RpoN binding profiles are essentially identical in samples obtained from cells grown in two different media. Second, and more importantly, we observe RpoN binding at the promoter upstream from *hrpL* in both conditions, only one of which results in *hrpL* transcription (as evidenced by the capture of a 5'-end). In many other cases, we observe RpoN binding unaccompanied by transcription.

Our preliminary survey identified 226 RpoN binding sites on the chromosome of DC3000, of which 214 have one or more associated RpoN motifs. The genes downstream from RpoN motifs encode proteins that are involved in multiple processes such as transport and binding proteins, metabolism, and regulation. Our RNA-Seq data detected 50 that are highly expressed (based on the "top 1000" transcription start site data set), among which 12 genes are differentially expressed 1.5 hours after a shift from a rich *hrp*-repressing medium to *hrp*-inducing medium. It is likely that longer culture times would have resulted in the expression of additional genes due to stress (nutrient depletion or other factors) such as those transcribed by RpoE or RpoH. It is worth noting that the "top 1000" data set excludes captured 5'-ends that are present at low levels. An examination of these might reveal the activation of additional genes.

The RpoN ChIP-Seq and RNA-Seq surveys were technically quite successful but should be considered only as initial steps in a thorough study of the RpoN regulon. Several additional experiments are described below for consideration.

Confirmation of candidate RpoN regulon members. Candidate promoters and their downstream targets should be re-evaluated using multiple methods including promoter traps, evidence of transcriptional dependence on RpoN, and mapping of transcription start sites. This will be especially important for promoters and genes that are suspected to have an important role in virulence.

One serious complication associated with RpoN promoter characterization (e.g., in a promoter trap experiment) is that transcriptional activity requires the presence of an enhancer binding site. The location of this site in most cases will not be known and could be very distant, 80 to 150 bp upstream or more (or even downstream) from the transcription start site [107]. Assuming an upstream location, nested deletions will probably be necessary to localize enhancer binding sites. A second challenge involves identifying the experimental conditions that are necessary to activate the EBPs. This is discussed in more detail in the next section.

Identification of conditions required for expression. Although the ChIP-Seq experiment may have identified most RpoN promoters in the DC3000 genome, transcription was monitored in only two conditions. In fact, only about 20 percent of potential RpoN-regulated genes were expressed in total. Without the means to "unlock" the remaining 80%, it will be difficult to study these genes in any transcriptional or functional experiments. The

difficulty is illustrated by HrpL, which is transcribed by RpoN when the bacteria are in *hrp*-inducing medium. Although it is known that the HrpR/S EBPs are required [70, 75], the precise environmental cue that activates the system is not understood. In other cases, the required condition can be attributed to the presence of a specific substrate. For example, *crcX* (regulated by RpoN) is required for growth of DC3000 in certain carbon sources such as mannitol and arabinose but not in succinate [217]. Some EBPs are controlled by RpoN itself, and presumably require other EBPs for activation. Other EBPs do not appear to have sensor-response domains, and thus may be functional when their protein levels reach an effective level.

It may be possible to explore a large number of conditions in parallel using a Biolog phenotype array coupled with appropriate promoter trap constructs and a reporter such as GFP. Such large-scale nutrient profiling, as described by Gardiner *et. al.,* is a promising approach [260]. A wide variety of individual promoters could be tested using this system. Alternatively, experiments could focus on EBPs rather than conditions. EBPs are relatively easy to identify and are limited in number in DC3000, compared to the potentially huge number of conditions that might need to be evaluated. It may be possible to "force" enhancer binding using genetic or other methods. It is feasible to engineer a promiscuous enhancer that globally activates RpoN-regulated genes as described by Samuels *et. al.* [108]. This offers some hope that it might be possible to target individual EBPs for activation.

Investigation of the relationship between effectors hopU1, *avrE, and RpoN.* RpoN binding and RpoN promoter motifs are observed upstream of *hopU1* and in the coding region of *avrE1*.

If RpoN controls transcription of either effector, it would be intriguing in two ways. First, genes in the Type III secretion system are thought to be controlled by HrpL exclusively [214] and no evidence has yet invalidated this model. Second, as explained below, the reasons why RpoN might regulate these genes are potentially interesting.

The RpoN binding site upstream of *hopU1* is located in the coding region towards the end of *hopF2* and is accompanied by a start site within 12 base pairs of the RpoN motif (59 counts in *hrpL-FLAG* 5' capture data). The closest *hrp* promoter is upstream of *shcF* and its neighbor *hopF2*, more than 1000 base pairs from the RpoN binding site (Figure 4.1). The RpoN promoter may help control the timing and level of transcription and expression of *hopU1,* whose function is to suppress the plant innate immune system [261-263].

The RpoN binding within *avrE1* is probably not positioned to regulate any downstream genes (Figure 4.2). However, it may reduce transcription of the last two thirds of *avrE1* by occluding RNA polymerase and consequently reduce the availability of this conserved effector. It may be important to note that AvrE1 has three domains, designated as WxxxE-1, WxxxE-2 and KK, which are required for the ability of AvrE1 to promote disease symptoms and suppress plant basal defenses [264]. The first domain, WxxE-1, unambiguously controlled by HrpL, is positioned upstream of the RpoN binding site, but RpoN might drive transcription for the WxxE-2 and KK domains downstream of its binding site when the appropriate enhancer is available. In our 5'-capture data (*hrp*-inducing, early time point), no transcription start site is observed near the RpoN motif. Still, it is possible that the promoter would be activated at a later time point *in vitro* or during infection.

Is the putative RpoN promoter within *avrE1* conserved in related bacterial strains? To look at this more closely, the ortholog sequences of *avrE1* among *Pseudomonadales*

were identified as reciprocal best BLAST "hits" using the methods described in chapter 2. The DNA sequences were then aligned and viewed in seaview [159]. The alignment in vicinity of RpoN motif is shown in Figure 4.3. The RpoN motif is commonly observed (66/75 sequences), suggesting that the motif may be functional—although the sequence surrounding the site is also well conserved. 40 bps downstream from the RpoN motif is a potential translational start codon (ATG). A protein starting from this codon will be in the same reading frame as the annotated full length AvrE1. We did not observe a Shine-Dalgarno sequence (ribosome binding site, RBS) [265] upstream from the ATG. The presence of an RBS is correlated with highly expressed genes [266]. However, genes without an RBS can also be expressed [267].

Whether this RpoN motif functions as a promoter requires further experimental evidence. It is also possible that RpoN binds to this site for reasons unrelated to transcription initiation. For example, bound RpoN could attenuate transcription from the HrpL promoter upstream. Future experiments could involve introducing silent mutations at the RpoN motif. These mutations would not change protein sequences but only the nucleotide sequence to abolish RpoN binding. If the RpoN motif is functional, one would expect to see changes in transcript ratio upstream and downstream from the site and perhaps even phenotypic changes in virulence. It should also be possible to detect a distinct transcript arising from the promoter. It may be advantageous to clone the distal part of *avrE1*, including the RpoN motif, in order to analyze the system in isolation.

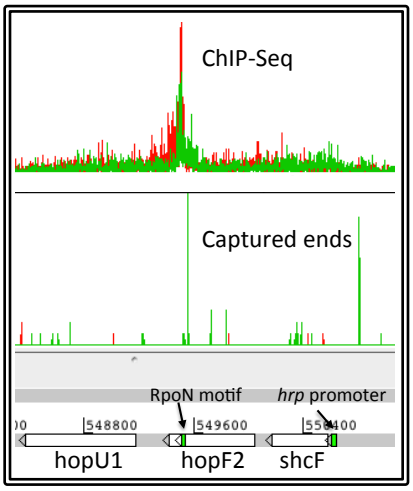


Figure 4. 1 RpoN motif near *hopU1*.

ChIP-Seq profile and 5' captured profiles are plotted above gene annotation. RpoN motif found in the coding region of *hopF2*. There is no *hrp* promoter directly upstream of *hopF2* but one is found upstream of *shcF*.

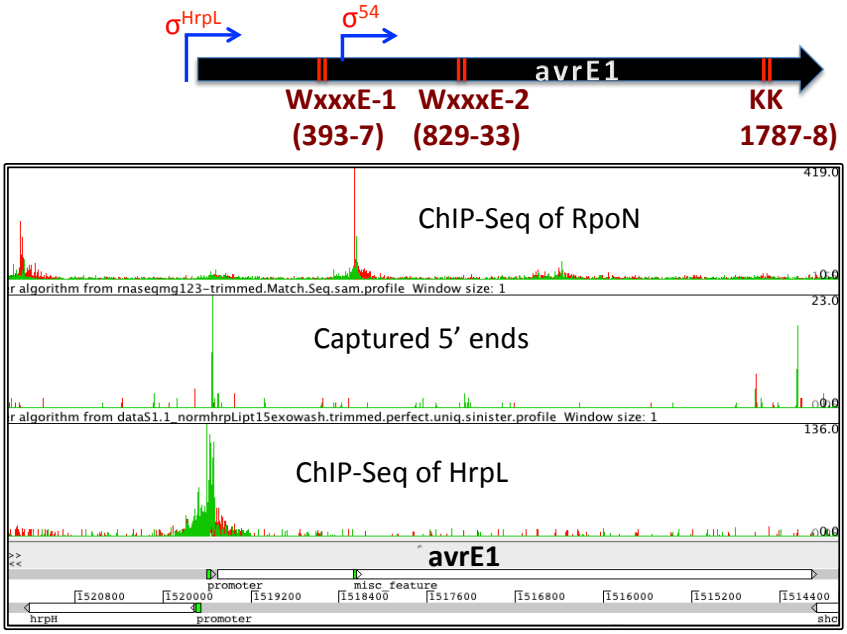


Figure 4. 2 RpoN motif near *avrE1*.

ChIP-Seq profile and 5' captured profiles are plotted above gene annotation. RpoN motif is in the intragenic region of *avrE1.*

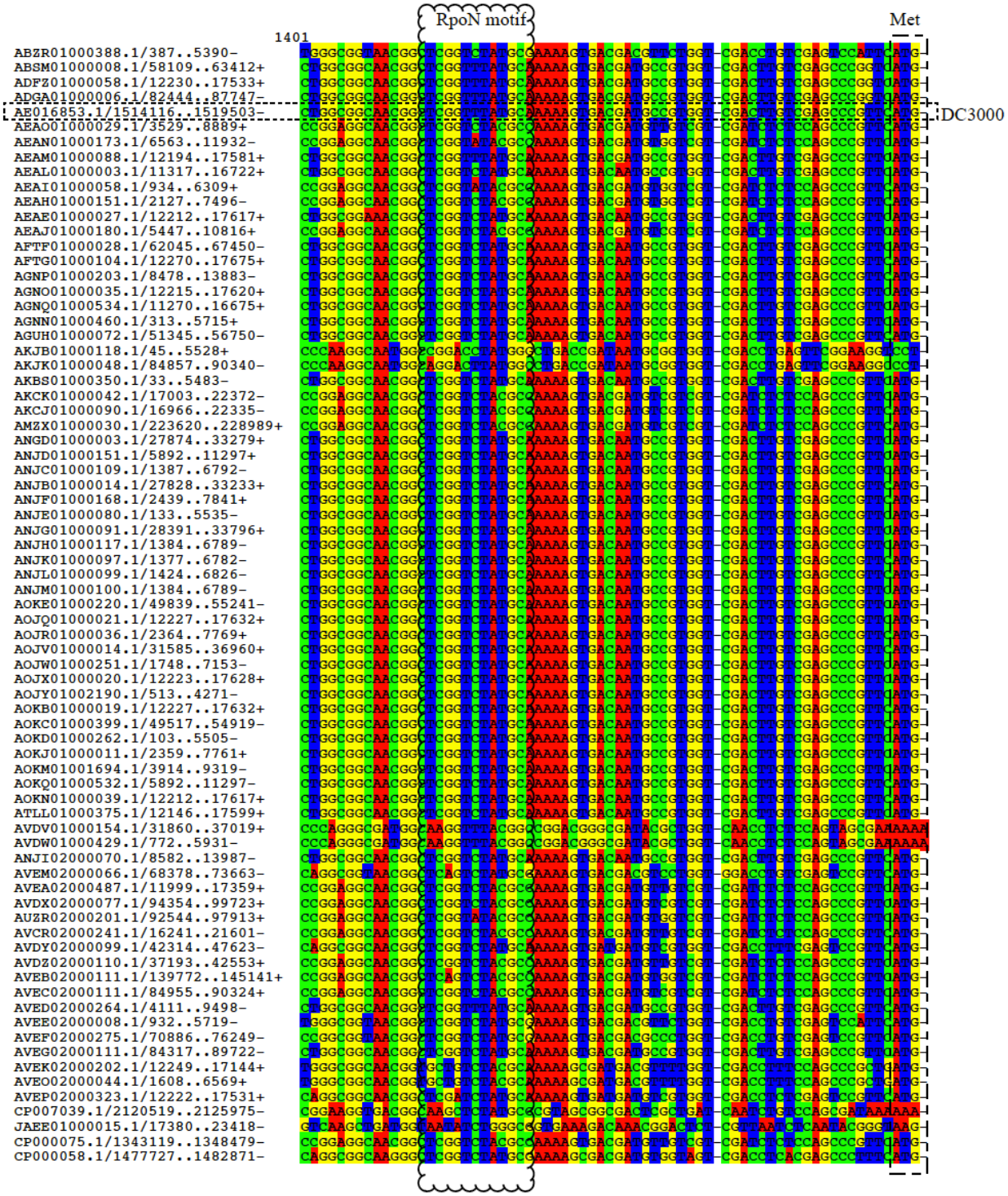

RpoN motif
Met
1401
DC3000

Figure 4. 3 Alignment of RpoN motif in coding region of *avrE1*

DNA sequence alignment of *avrE1* orthologs at the region surrounding the RpoN binding site is shown. Orthologs were defined by reciprocal best BLAST as described in chapter 2. Alignment and visualization were accomplished using SeaView [159]. Labels to the left of the sequence alignment are constructed using the Accession number of the genome, followed by the position of the *avrE1* ortholog in the genome and its chromosomal strand.

Investigation of genes that are differentially expressed in hrp-inducing medium. It will be very important to explore any regulatory genes in the RpoN regulon, particularly those that are expressed in parallel with *hrpL*. These genes may lead to the identification of new virulence factors—functions related to infection, or genes involved in supporting roles (for example, harvesting plant resources). PSPTO_1255 and PSPTO_4171, both encoding amino acid ABC transporter, periplasmic amino acid-binding proteins, may be involved in nutrient uptake. They are upregulated in *hrp*-inducing conditions and thus may be important during infection. On the contrary, PSPTO_1277, encoding for glycine cleavage system H protein, is down regulated in *hrp*-inducing conditions. The glycine cleavage system in bacteria can be used for glycine synthesis [268]. The bacteria may need to reduce glycine synthesis once glycine is available from the host. Overall, DC3000 may coordinate gene expression in the T3SS and other processes for the benefit of its growth *in planta*.

Table 4. 2 Sequences identified by IVET that may contain *hrp* promoters or RpoN promoters

Genes and function of genes downstream of the IVET locus [269] are listed.
RpoN binding site: "no/weak" for regions that show ChIP-Seq peaks below cutoff used in our study and thus there is no RpoN found. "yes" for good RpoN binding sites.
Genes likely to be regulated by RpoN are in red.

Gene	Function	IVET locus	IVET sequence	RpoN binding site	RpoN motif	HrpL binding site/motif
1404	hrpL	ipx1	1542097..1542626	yes	yes	no
1369	shcN	ipx15	complement(1504698..1505257)	no	no	yes
1387	hrpF	ipx2	complement(1528033..1528534)	no	no	yes
1381	hrpA1	ipx4	complement(1524078..1524583)	no	no	yes
524	M20/M25/M40 family peptidase	ipx41	complement(572065..572551)	no	no	yes
317	glycine cleavage system protein H	ipx44	complement(344748..345264)	no/weak	no	no
2105	thiamine biosynthesis lipoprotein, putative	ipx45	complement(2279738..2280310)	no	no	yes
370	MATE efflux family protein	ipx54	complement(404677..405132)	no	no	yes
1529	arsC family protein	ipx58	1686879..1687519	no/weak	no	no
1738	hypothetical protein	ipx62	complement(1906222..1906801)	yes	yes	no
3087	hopAB2	ipx7	3470043..3470591	no	no	yes
357	hopS1	ipx70/ ipx69	5191701..5192181/ 5191827..5192368	no/weak	no	no
192	recombinase, putative	ipx73	215296..215847	yes	no	no
1022	hopAM1-1	ipx8	1116001..1116405	no	no	yes
4727	hopD	ipx9	complement(5349489..5350069)	no	no	yes
1377	avrE1	ipx10	complement(1519747..1520339)	no	no	yes
2059	hypothetical protein	ipx64	complement(2239067..2239547)	yes	yes	no

Investigation of genes whose upstream sequences are induced in in planta conditions. One DC3000 genomic data set that has often been overlooked consists of chromosomal regions isolated using *in vivo* expression technology (IVET) in Barbara Kunkel's laboratory [269]. These regions were identified as showing promoter activity in *Arabidopsis* during infection because they were able to drive transcription of a promoterless *hrpC* gene (this gene is essential for disease). In Table 4.2 the captured regions are compared with the ChIP-Seq profiles generated using *rpoN-FLAG* and *hrpL-FLAG*. IVET regions that overlap putative RpoN or HrpL binding sites are especially worthy of further investigation to determine if infection-related promoters are be embedded in them. Two genes of particular interest are PSPTO_1738 and PSPTO_2059, both of which encode hypothetical proteins. The RpoN binding sites and motifs in their upstream regions suggest that they are regulated by RpoN. A pfam search [270] of PSPTO_1738 retrieves a TENA/THI-4/PQQC family domain. Genes having this domain enhance the expression of extracellular enzymes in *Bacillus subtilis* [271], are involved in thiamine biosynthesis in *Neurospora crassa* [272], and Pyrroloquinoline quinone synthesis in *Methylobacterium extorquens* [273]. PSPTO_2059 has an ATP-grasp superfamily pfam domain that is found in carboxylate-amine/thiol ligases [274]. It is not clear what the function of the two genes are but the fact that they can be regulated by RpoN and induced in the plant host during infection is extremely interesting.

REFERENCES

1. Ait Tayeb, L., et al., *Molecular phylogeny of the genus Pseudomonas based on rpoB sequences and application for the identification of isolates.* Research in Microbiology, 2005. **156**(5-6): p. 763-73.
2. Morris, C.E., et al., *Surprising niche for the plant pathogen Pseudomonas syringae.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2007. **7**(1): p. 84-92.
3. Morris, C.E., et al., *The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle.* The ISME journal, 2008. **2**(3): p. 321-34.
4. Bradbury, J., *Guide to plant pathogenic bacteria.* CAB International Mycological Institute (CMI), UK, 1986. **332**.
5. Nikel, P.I., et al., *The private life of environmental bacteria: pollutant biodegradation at the single cell level.* Environmental Microbiology, 2014. **16**(3): p. 628-642.
6. Samuel, M.S., A. Sivaramakrishna, and A. Mehta, *Bioremediation of p-Nitrophenol by Pseudomonas putida 1274 strain.* Journal of environmental health science & engineering, 2014. **12**(1): p. 53.
7. Paliwal, V., et al., *Pseudomonas putida CSV86: a candidate genome for genetic bioaugmentation.* Plos One, 2014. **9**(1): p. e84000.
8. Nitisakulkan, T., et al., *Degradation of chloroanilines by toluene dioxygenase from Pseudomonas putida T57.* Journal of bioscience and bioengineering, 2014. **117**(3): p. 292-7.
9. Sajeli Begum, A., et al., *Isolation and Characterization of Antimicrobial Cyclic Dipeptides from Pseudomonas fluorescens and Their Efficacy on Sorghum Grain Mold Fungi.* Chemistry & Biodiversity, 2014. **11**(1): p. 92-100.
10. Barton, M.D., et al., *The genome of Pseudomonas fluorescens strain R124 demonstrates phenotypic adaptation to the mineral environment.* Journal of bacteriology, 2013. **195**(21): p. 4793-803.
11. Tummler, B., et al., *Advances in understanding Pseudomonas.* F1000prime reports, 2014. **6**: p. 9.
12. Volko, S.M., T. Boller, and F.M. Ausubel, *Isolation of new Arabidopsis mutants with enhanced disease susceptibility to Pseudomonas syringae by direct screening.* Genetics, 1998. **149**(2): p. 537-48.
13. Whalen Maureen C, I.R.W., Bent Andrew F, Staskawicz Brian J, *ldentification of Pseudomonas syringae Pathogens of Arabidopsis and a Bacterial Locus Determining Avirulence on Both Arabidopsis and Soybean* The Plant Cell, 1991. **3**: p. 49-59.
14. Ainsworth, D.A.C.a.T., *Molecular and Physiological Characterization of Pseudomonas syringae pv. tomato and Pseudomonas yringae pv. maculicola Strains That Produce the Phytotoxin Coronatine.* Applied and Environmental Microbiology, 1995. **61**(10): p. 3530-3536.
15. Maki, L.R., et al., *Ice nucleation induced by pseudomonas syringae.* Applied microbiology, 1974. **28**(3): p. 456-9.

16. Sesartic, A., U. Lohmann, and T. and Storelvmo, *Bacteria in the ECHAM5-HAM global climate model.* Atmos. Chem. Phys. Discuss., 2011. **11**: p. 32.
17. Nishimori, E., K. Kita-Tsukamoto, and H. Wakabayashi, *Pseudomonas plecoglossicida sp. nov., the causative agent of bacterial haemorrhagic ascites of ayu, Plecoglossus altivelis.* International journal of systematic and evolutionary microbiology, 2000. **50 Pt 1**: p. 83-9.
18. Park, S.C., et al., *Isolation of bacteriophages specific to a fish pathogen, Pseudomonas plecoglossicida, as a candidate for disease control.* Applied and environmental microbiology, 2000. **66**(4): p. 1416-22.
19. Suzuki, M., et al., *Genome Sequence of a Strain of the Human Pathogenic Bacterium Pseudomonas alcaligenes That Caused Bloodstream Infection.* Genome announcements, 2013. **1**(5).
20. Salch, Y.P. and P.D. Shaw, *Isolation and characterization of pathogenicity genes of Pseudomonas syringae pv. tabaci.* Journal of bacteriology, 1988. **170**(6): p. 2584-91.
21. Reinhardt, J.A., et al., *De novo assembly using low-coverage short read sequence data from the rice pathogen Pseudomonas syringae pv. oryzae.* Genome Research, 2009. **19**(2): p. 294-305.
22. Cheol-In Kang, et al., *Pseudomonas aeruginosa Bacteremia: Risk Factors for Mortality and Influence of Delayed Receipt of Effective Antimicrobial Therapy on Clinical Outcome.* Clinical Infectious Diseases, 2003. **37**(6): p. 7.
23. Rahme, L.G., et al., *Plants and animals share functionally common bacterial virulence factors.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(16): p. 8815-21.
24. Lindholm, D.A., et al., *Novel Pseudomonas fluorescens septic sacroiliitis in a healthy soldier.* Military medicine, 2013. **178**(8): p. e963-6.
25. Morris, C.E., C.L. Monteil, and O. Berge, *The life history of Pseudomonas syringae: linking agriculture to earth system processes.* Annual review of phytopathology, 2013. **51**: p. 85-104.
26. Hwang, M.S., et al., *Phylogenetic characterization of virulence and resistance phenotypes of Pseudomonas syringae.* Applied and environmental microbiology, 2005. **71**(9): p. 5182-91.
27. Underwood, W., M. Melotto, and S.Y. He, *Role of plant stomata in bacterial invasion.* Cellular Microbiology, 2007. **9**(7): p. 1621-9.
28. Melotto, M., et al., *Plant stomata function in innate immunity against bacterial invasion.* Cell, 2006. **126**(5): p. 969-80.
29. Lee, S., et al., *Coronatine inhibits stomatal closure and delays hypersensitive response cell death induced by nonhost bacterial pathogens.* PeerJ, 2013. **1**: p. e34.
30. Zheng, X.-y., et al., *Coronatine Promotes Pseudomonas syringae Virulence in Plants by Activating a Signaling Cascade that Inhibits Salicylic Acid Accumulation.* Cell Host & Microbe, 2012. **11**(6): p. 587-596.
31. Xin, X.F. and S.Y. He, *Pseudomonas syringae pv. tomato DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants.* Annual review of phytopathology, 2013. **51**: p. 473-98.
32. Lawton, K., et al., *Systemic acquired resistance in Arabidopsis requires salicylic acid but not ethylene.* Molecular plant-microbe interactions : MPMI, 1995. **8**(6): p. 863-70.

33. Gaffney, T., et al., *Requirement of salicylic Acid for the induction of systemic acquired resistance.* Science, 1993. **261**(5122): p. 754-6.
34. Rico, A. and G.M. Preston, *Pseudomonas syringae pv. tomato DC3000 uses constitutive and apoplast-induced nutrient assimilation pathways to catabolize nutrients that are abundant in the tomato apoplast.* 2008(0894-0282 (Print)).
35. Wei, H.L. and A. Collmer, *Multiple lessons from the multiple functions of a regulator of type III secretion system assembly in the plant pathogen Pseudomonas syringae.* Molecular Microbiology, 2012. **85**(2): p. 195-200.
36. Wei, H.L., et al., *Consequences of flagellin export through the type III secretion system of Pseudomonas syringae reveal a major difference in the innate immune systems of mammals and the model plant Nicotiana benthamiana.* Cellular Microbiology, 2012.
37. Lozano-Duran, R., et al., *The bacterial effector HopM1 suppresses PAMP-triggered oxidative burst and stomatal immunity.* The New phytologist, 2014. **202**(1): p. 259-69.
38. Block, A., et al., *The Pseudomonas syringae type III effector HopD1 suppresses effector-triggered immunity, localizes to the endoplasmic reticulum, and targets the Arabidopsis transcription factor NTL9.* The New phytologist, 2014. **201**(4): p. 1358-70.
39. Raymond, B., et al., *Subversion of trafficking, apoptosis, and innate immunity by type III secretion system effectors.* Trends in microbiology, 2013. **21**(8): p. 430-41.
40. Jiang, S., et al., *Bacterial effector activates jasmonate signaling by directly targeting JAZ transcriptional repressors.* PLoS pathogens, 2013. **9**(10): p. e1003715.
41. Chatterjee, S., et al., *Structure and biophysics of type III secretion in bacteria.* Biochemistry, 2013. **52**(15): p. 2508-17.
42. Abrusci, P., et al., *Building a secreting nanomachine: a structural overview of the T3SS.* Current opinion in structural biology, 2014. **25C**: p. 111-117.
43. Kawamoto, A., et al., *Common and distinct structural features of Salmonella injectisome and flagellar basal body.* Scientific reports, 2013. **3**: p. 3369.
44. Alfano, J.R., et al., *The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(9): p. 4856-4861.
45. Sawada, H., et al., *Phylogenetic analysis of Pseudomonas syringae pathovars suggests the horizontal gene transfer of argK and the evolutionary stability of hrp gene cluster.* Journal of molecular evolution, 1999. **49**(5): p. 627-44.
46. Pitman, A.R., et al., *Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants.* Current biology : CB, 2005. **15**(24): p. 2230-5.
47. Buell, C.R., et al., *The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(18): p. 10181-6.
48. Lindeberg, M., et al., *Roadmap to new virulence determinants in Pseudomonas syringae: Insights from comparative genomics and genome organization.* Molecular Plant-Microbe Interactions, 2008. **21**(6): p. 685-700.
49. Chatterjee, A., et al., *GacA, the response regulator of a two-component system, acts as a master regulator in Pseudomonas syringae pv. tomato DC3000 by controlling*

regulatory RNA, transcriptional activators, and alternate sigma factors. Molecular plant-microbe interactions : MPMI, 2003. **16**(12): p. 1106-17.
50. Ortiz-Martín, I., et al., *Positive Regulation of the Hrp Type III Secretion System in Pseudomonas syringae pv. phaseolicola.* MPMI, 2010. **23**(5): p. 7.
51. David K. Willis, E.M.H., Jessica J. Rich, Terese M. Barta, Steven E. Lindow, and Nickolas J. Panopoulos, *Isolation and Characterization of a Pseudomonas syringae pv. syringae Mutant Deficient in Lesion Formation on Bean.* Molecular Plant-Microbe Interactions, 1990. **3**(3): p. 7.
52. Diallo, M.D., et al., *Pseudomonas syringae naturally lacking the canonical type III secretion system are ubiquitous in nonagricultural habitats, are phylogenetically diverse and can be pathogenic.* Isme Journal, 2012. **6**(7): p. 1325-1335.
53. Tampakaki, A.P., *Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria.* Frontiers in plant science, 2014. **5**: p. 114.
54. Coombes, B.K., *Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria.* Trends in microbiology, 2009. **17**(3): p. 89-94.
55. Finn, R.D., et al., *Escherichia coli RNA polymerase core and holoenzyme structures.* The EMBO journal, 2000. **19**(24): p. 6833-44.
56. Wagner, R., *Transcription Regulation in Prokaryotes*, 2000, Oxford University Press. p. 380.
57. Helmann, J.D., *The extracytoplasmic function (ECF) sigma factors*, in *Advances in Microbial Physiology*2002, Academic Press. p. 47-110.
58. Kazmierczak, M.J., M. Wiedmann, and K.J. Boor, *Alternative Sigma Factors and Their Roles in Bacterial Virulence.* Microbiology and Molecular Biology Reviews, 2005. **69**(4): p. 527-543.
59. Paget, M. and J. Helmann, *The sigma70 family of sigma factors.* Genome Biology, 2003. **4**(1): p. 6.
60. Markel, E., et al., *Regulons of three Pseudomonas syringae pv. tomato DC3000 iron starvation sigma factors.* Applied and environmental microbiology, 2013. **79**(2): p. 725-7.
61. Oguiza, J.A., K. Kiil, and D.W. Ussery, *Extracytoplasmic function sigma factors in Pseudomonas syringae.* Trends in microbiology, 2005. **13**(12): p. 565-8.
62. Filiatrault, M.J., et al., *Genome-wide identification of transcriptional start sites in the plant pathogen Pseudomonas syringae pv. tomato str. DC3000.* PLoS One, 2011. **6**(12): p. e29335.
63. Swingle, B., et al., *Characterization of the PvdS-regulated promoter motif in Pseudomonas syringae pv. tomato DC3000 reveals regulon members and insights regarding PvdS function in other pseudomonads.* Molecular Microbiology, 2008. **68**(4): p. 871-89.
64. Markel, E., et al., *An extracytoplasmic function sigma factor-mediated cell surface signaling system.* J. Bacteriol., 2011. **193**(20): p. 9.
65. Ferreira, A.O., et al., *Whole-genome expression profiling defines the HrpL regulon of Pseudomonas syringae pv. tomato DC3000, allows de novo reconstruction of the Hrp cis clement, and identifies novel coregulated genes.* Molecular plant-microbe interactions : MPMI, 2006. **19**(11): p. 1167-79.

66. Heeb, S. and D. Haas, *Regulatory roles of the GacS/GacA two-component system in plant-associated and other gram-negative bacteria.* Molecular plant-microbe interactions : MPMI, 2001. **14**(12): p. 1351-63.
67. Kitten, T., et al., *A newly identified regulator is required for virulence and toxin production in.* Mol Microbiol., 1998. **28**(5): p. 11.
68. Hendrickson, E.L., P. Guevera, and F.M. Ausubel, *The alternative sigma factor RpoN is required for hrp activity in Pseudomonas syringae pv. maculicola and acts at the level of hrpL transcription.* Journal of bacteriology, 2000. **182**(12): p. 3508-16.
69. Schuster, M. and C. Grimm, *Domain switching betweenc hrpR and hrpS affects the regulatory function of the hybrid genes in Pseudomonas syringae pv. phaseolicola.* Molecular plant pathology, 2000. **1**(4): p. 9.
70. Hutcheson, S.W., et al., *Enhancer-Binding Proteins HrpR and HrpS Interact To Regulate hrp-Encoded Type III Protein Secretion in Pseudomonas syringae Strains.* J Bacteriol, 2001. **183**(19): p. 5589-5598.
71. Xiao, Y., et al., *Identification of a putative alternate sigma factor and characterization of a multicomponent regulatory cascade controlling the expression of Pseudomonas syringae pv. syringae Pss61 hrp and hrmA genes.* Journal of bacteriology, 1994. **176**(4): p. 1025-36.
72. Tang, X.Y., Y.M. Xiao, and J.M. Zhou, *Regulation of the type III secretion system in phytopathogenic bacteria.* Molecular Plant-Microbe Interactions, 2006. **19**(11): p. 1159-1166.
73. Wei, W., et al., *The gene coding for the Hrp pilus structural protein is required for type III secretion of Hrp and Avr proteins in Pseudomonas syringae pv. tomato.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(5): p. 2247-52.
74. Ortiz-Martin, I., et al., *Negative Regulation of the Hrp Type III Secretion System in Pseudomonas syringae pv. phaseolicola.* Molecular Plant-Microbe Interactions, 2010. **23**(5): p. 682-701.
75. Jovanovic, M., et al., *Regulation of the co-evolved HrpR and HrpS AAA+ proteins required for Pseudomonas syringae pathogenicity.* Nat Commun, 2011. **2**: p. 177.
76. Staskawicz, B.J., D. Dahlbeck, and N.T. Keen, *Cloned avirulence gene of Pseudomonas syringae pv. glycinea determines race-specific incompatibility on Glycine max (L.) Merr.* Proceedings of the National Academy of Sciences of the United States of America, 1984. **81**(19): p. 6024-8.
77. Kobayashi, D.Y., S.J. Tamaki, and N.T. Keen, *Cloned avirulence genes from the tomato pathogen Pseudomonas syringae pv. tomato confer cultivar specificity on soybean.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(1): p. 157-61.
78. Shen, H. and N.T. Keen, *Characterization of the promoter of avirulence gene D from Pseudomonas syringae pv. tomato.* Journal of bacteriology, 1993. **175**(18): p. 5916-24.
79. Collmer, A., et al., *Genomic mining type III secretion system effectors in Pseudomonas syringae yields new picks for all TTSS prospectors.* Trends in microbiology, 2002. **10**(10): p. 462-9.
80. Fouts, D.E., et al., *Genomewide identification of Pseudomonas syringae pv. tomato DC3000 promoters controlled by the HrpL alternative sigma factor.* Proceedings of

the National Academy of Sciences of the United States of America, 2002. **99**(4): p. 2275-80.
81. Guttman, D.S., et al., *A functional screen for the type III (Hrp) secretome of the plant pathogen Pseudomonas syringae.* Science, 2002. **295**(5560): p. 1722-1726.
82. Julie Zwiesler-Vollick, A.E.P.-J., Kinya Nomura, Sruti Bandyopadhyay, Vinita Joardar, Barbara N. Kunkel and Sheng Yang He, *Identification of novel hrp-regulated genes through functional genomic analysis of the Pseudomonas syringae pv. tomato DC3000 genome.* Molecular Microbiology, 2002. **45**(5): p. 11.
83. Petnicki-Ocwieja, T., et al., *Genomewide identification of proteins secreted by the Hrp type III protein secretion system of Pseudomonas syringae pv. tomato DC3000.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(11): p. 7652-7.
84. Greenberg, J., *Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells.* Current Opinion in Microbiology, 2003. **6**(1): p. 20-28.
85. Chang, J.H., et al., *A high-throughput, near-saturating screen for type III effector genes from Pseudomonas syringae.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(7): p. 2549-54.
86. Magdalen Lindeberg, S.C., Christopher R. Myers, Lisa M. Schechter, David J. Schneider, and Alan Collmer, *Closing the Circle on the Discovery of Genes Encoding Hrp Regulon Members and Type III Secretion System Effectors in the Genomes of Three Model Pseudomonas syringae Strains* MPMI, 2006. **19**(11): p. 8.
87. Schechter, L.M., et al., *Multiple approaches to a complete inventory of Pseudomonas syringae pv. tomato DC3000 type III secretion system effector proteins.* Molecular plant-microbe interactions : MPMI, 2006. **19**(11): p. 1180-92.
88. Vencato, M., et al., *Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of Pseudomonas syringae pv. phaseolicola 1448A.* Molecular Plant-Microbe Interactions, 2006. **19**(11): p. 1193-1206.
89. Kvitko, B.H., et al., *Deletions in the Repertoire of Pseudomonassyringae pv. tomato DC3000 Type III Secretion Effector Genes Reveal Functional Overlap among Effectors*, in *Plos Pathogens*2009. p. 1-16.
90. Cunnac, S., et al., *Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in Pseudomonas syringae.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(7): p. 2975-80.
91. Merrick, M.J. and J.R. Coppard, *Mutations in genes downstream of the rpoN gene (encoding sigma 54) of Klebsiella pneumoniae affect expression from sigma 54-dependent promoters.* Molecular Microbiology, 1989. **3**(12): p. 1765-75.
92. Studholme, D.J. and M. Buck, *The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences.* Fems Microbiology Letters, 2000. **186**(1): p. 1-9.
93. Merrick, M.J., *In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N).* Molecular Microbiology, 1993. **10**(5): p. 903-9.
94. Buck, M. and W. Cannon, *Specific binding of the transcription factor sigma-54 to promoter DNA.* Nature, 1992. **358**(6385): p. 422-4.
95. Dombroski, A.J., W.A. Walter, and C.A. Gross, *Amino-terminal amino acids modulate sigma-factor DNA-binding activity.* Genes & development, 1993. **7**(12A): p. 2446-55.

96. Dombroski, A.J., et al., *Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA.* Cell, 1992. **70**(3): p. 501-12.
97. Sasse-Dwight, S. and J.D. Gralla, *Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma 54.* (0092-8674 (Print)).
98. Rombel, I., et al., *The bacterial enhancer-binding protein NtrC as a molecular machine.* Cold Spring Harbor symposia on quantitative biology, 1998. **63**: p. 157-66.
99. Buck, M., et al., *Upstream activator sequences are present in the promoters of nitrogen fixation genes.* Nature, 1986. **320**(6060): p. 374-378.
100. Reitzer, L.J. and B. Magasanik, *Transcription of glnA in E. coli is stimulated by activator bound to sites far from the promoter.* Cell. **45**(6): p. 785-792.
101. Beck, L.L., T.G. Smith, and T.R. Hoover, *Look, no hands! Unconventional transcriptional activators in bacteria.* Trends in microbiology, 2007. **15**(12): p. 530-7.
102. Batchelor, J.D., et al., *Structural mechanism of GAF-regulated sigma(54) activators from Aquifex aeolicus.* Journal of Molecular Biology, 2013. **425**(1): p. 156-70.
103. Barrios, H., B. Valderrama, and E. Morett, *Compilation and analysis of sigma(54)-dependent promoter sequences.* Nucleic Acids Research, 1999. **27**(22): p. 4305-13.
104. Rendon, M.A., et al., *Sigma factor RpoN (sigma54) regulates pilE transcription in commensal Neisseria elongata.* Molecular Microbiology, 2013. **90**(1): p. 103-13.
105. Vidangos, N., et al., *Structure, function, and tethering of DNA-binding domains in sigma(54) transcriptional activators.* Biopolymers, 2013. **99**(12): p. 1082-96.
106. Sysoeva, T.A., et al., *Crystallization and preliminary X-ray analysis of the ATPase domain of the sigma(54)-dependent transcription activator NtrC1 from Aquifex aeolicus bound to the ATP analog ADP-BeFx.* Acta crystallographica. Section F, Structural biology and crystallization communications, 2013. **69**(Pt 12): p. 1384-8.
107. Bush, M. and R. Dixon, *The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription.* Microbiology and molecular biology reviews : MMBR, 2012. **76**(3): p. 497-529.
108. Samuels, D.J., et al., *Use of a promiscuous, constitutively-active bacterial enhancer-binding protein to define the sigma(5)(4) (RpoN) regulon of Salmonella Typhimurium LT2.* BMC genomics, 2013. **14**: p. 602.
109. Qiu, Y., et al., *Characterizing the interplay between multiple levels of organization within bacterial sigma factor regulatory networks.* Nature Communications, 2013. **4**.
110. Sana, T.G., et al., *Divergent Control of Two Type VI Secretion Systems by RpoN in Pseudomonas aeruginosa.* Plos One, 2013. **8**(10): p. e76030.
111. Ramos, L.S., et al., *The fire blight pathogen Erwinia amylovora requires the rpoN gene for pathogenicity in apple.* Molecular plant pathology, 2013. **14**(8): p. 838-43.
112. Hao, B., et al., *Role of alternative sigma factor 54 (RpoN) from Vibrio anguillarum M3 in protease secretion, exopolysaccharide production, biofilm formation, and virulence.* Applied microbiology and biotechnology, 2013. **97**(6): p. 2575-85.
113. Leang, C., et al., *Genome-wide analysis of the RpoN regulon in Geobacter sulfurreducens.* BMC genomics, 2009. **10**: p. 331.
114. Salazar, E., et al., *Characterization of the NifA-RpoN Regulon in Rhizobium etli in Free Life and in Symbiosis with Phaseolus vulgaris.* Applied and environmental microbiology, 2010. **76**(13): p. 4510-4520.

115. Dong, T.G. and J.J. Mekalanos, *Characterization of the RpoN regulon reveals differential regulation of T6SS and new flagellar operons in Vibrio cholerae O37 strain V52.* Nucleic Acids Research, 2012. **40**(16): p. 7766-7775.
116. Sullivan, J.T., S.D. Brown, and C.W. Ronson, *The NifA-RpoN regulon of Mesorhizobium loti strain R7A and its symbiotic activation by a novel LacI/GalR-family regulator.* Plos One, 2013. **8**(1): p. e53762.
117. Damron, F.H., et al., *Analysis of the Pseudomonas aeruginosa regulon controlled by the sensor kinase KinB and sigma factor RpoN.* Journal of bacteriology, 2012. **194**(6): p. 1317-30.
118. Wei, C.F., et al., *A Pseudomonas syringae pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant Nicotiana benthamiana.* The Plant journal : for cell and molecular biology, 2007. **51**(1): p. 32-46.
119. Jin, Q.L., et al., *Type III protein secretion in Pseudomonas syringae.* Microbes and Infection, 2003. **5**(4): p. 301-310.
120. Alfano, J.R. and A. Collmer, *The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death.* Journal of bacteriology, 1997. **179**(18): p. 5655-62.
121. Lindeberg, M., et al., *Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen Pseudomonas syringae.* Molecular plant-microbe interactions : MPMI, 2005. **18**(4): p. 275-82.
122. Deslandes, L. and S. Rivas, *Catch me if you can: bacterial effectors and plant targets.* Trends in plant science, 2012. **17**(11): p. 644-55.
123. Mudgett, M.B., *New insights to the function of phytopathogenic bacterial type III effectors in plants.* Annual review of plant biology, 2005. **56**: p. 509-31.
124. Feng, F. and J.M. Zhou, *Plant-bacterial pathogen interactions mediated by type III effectors.* Current opinion in plant biology, 2012. **15**(4): p. 469-76.
125. Abramovitch, R.B., J.C. Anderson, and G.B. Martin, *Bacterial elicitation and evasion of plant innate immunity.* Nature reviews. Molecular cell biology, 2006. **7**(8): p. 601-11.
126. Kim, J.F. and J.R. Alfano, *Pathogenicity islands and virulence plasmids of bacterial plant pathogens.* Current topics in microbiology and immunology, 2002. **264**(2): p. 127-47.
127. Yuan, J. and S.Y. He, *The Pseudomonas syringae Hrp regulation and secretion system controls the production and secretion of multiple extracellular proteins.* Journal of bacteriology, 1996. **178**(21): p. 6399-402.
128. Xiao, Y. and S.W. Hutcheson, *A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in Pseudomonas syringae.* Journal of bacteriology, 1994. **176**(10): p. 3089-91.
129. Buell, C.R., et al., *The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(18): p. 10181-10186.
130. Mucyn, T.S., et al., *Variable suites of non-effector genes are co-regulated in the type III secretion virulence regulon across the Pseudomonas syringae phylogeny.* PLoS pathogens, 2014. **10**(1): p. e1003807.

131. Filiatrault, M.J., et al., *Transcriptome analysis of Pseudomonas syringae identifies new genes, noncoding RNAs, and antisense activity.* Journal of bacteriology, 2010. **192**(9): p. 2359-72.
132. Rhee, H.S. and B.F. Pugh, *ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy.* Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], 2012. **Chapter 21**: p. Unit 21 24.
133. Lefu Lan, X.D., Jianmin Zhou, and Xiaoyan Tang, *Genome-Wide Gene Expression Analysis of Pseudomonas syringae pv. tomato DC3000 Reveals Overlapping and Distinct Pathways Regulated by hrpL and hrpRS.* MPMI, 2006. **19**(9): p. 10.
134. Bronstein, P.A., et al., *Global transcriptional responses of Pseudomonas syringae DC3000 to changes in iron bioavailability in vitro.* Bmc Microbiology, 2008. **8**: p. 209.
135. King, E.O., M.K. Ward, and D.E. Raney, *Two simple media for the demonstration of pyocyanin and fluorescin.* The Journal of laboratory and clinical medicine, 1954. **44**(2): p. 301-7.
136. Huynh, T.V., D. Dahlbeck, and B.J. Staskawicz, *Bacterial blight of soybean: regulation of a pathogen gene determining host cultivar specificity.* Science, 1989. **245**(4924): p. 1374-7.
137. Hanahan, D., *Techniques for transformation of E. coli*, in *DNA cloning: a practical approach*1985, IRL Press: Oxford, United Kingdom. p. 26.
138. Schaefer, A., et al., *Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum.* Gene, 1994. **145**(1): p. 69-73.
139. Oh, H.S., et al., *Pseudomonas syringae lytic transglycosylases coregulated with the type III secretion system contribute to the translocation of effector proteins into plant cells.* Journal of bacteriology, 2007. **189**(22): p. 8277-8289.
140. DeLisa, M.P., et al., *Genetic analysis of the twin arginine translocator secretion pathway in bacteria.* J Biol Chem, 2002. **277**(33): p. 29825-31.
141. Buell, C.R., et al., *The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000*, in *PNAS*2003. p. 1-6.
142. Markel, E., et al., *An extracytoplasmic function sigma factor-mediated cell surface signaling system in Pseudomonas syringae pv. tomato DC3000 regulates gene expression in response to heterologous siderophores.* Journal of bacteriology, 2011. **193**(20): p. 5775-83.
143. Badel, J.L., et al., *A Pseudomonas syringae pv. tomato avrE1/hopM1 mutant is severely reduced in growth and lesion formation in tomato.* Molecular Plant-Microbe Interactions, 2006. **19**(2): p. 99-111.
144. Bronstein, P.A., et al., *Identification of a twin-arginine translocation system in Pseudomonas syringae pv. tomato DC3000 and its contribution to pathogenicity and fitness.* Journal of bacteriology, 2005. **187**(24): p. 8450-61.
145. Choi, K.H., A. Kumar, and H.P. Schweizer, *A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid transformation.* Journal of Microbiological Methods, 2006. **64**(3): p. 391-397.

146. Kvitko, B.H., et al., *Identification of harpins in Pseudomonas syringae pv. tomato DC3000, which are functionally similar to HrpK1 in promoting translocation of type III secretion system effectors.* Journal of Bacteriology, 2007. **189**(22): p. 8059-72.
147. Horton, R.M., et al., *Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.* Gene, 1989. **77**(1): p. 61-8.
148. Rhee, H.S. and B.F. Pugh, *Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution.* Cell, 2011. **147**(6): p. 1408-19.
149. Butcher, B.G., et al., *Characterization of the Fur regulon in Pseudomonas syringae pv. tomato DC3000.* Journal of bacteriology, 2011. **193**(18): p. 4598-611.
150. Patel, R.K. and M. Jain, *NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data.* Plos One, 2012. **7**(2).
151. Li, R.Q., et al., *SOAP2: an improved ultrafast tool for short read alignment.* Bioinformatics, 2009. **25**(15): p. 1966-1967.
152. Rutherford, K., et al., *Artemis: sequence visualization and annotation.* Bioinformatics, 2000. **16**(10): p. 944-5.
153. Albert, I., et al., *GeneTrack - a genomic data processing and visualization framework.* Bioinformatics, 2008. **24**(10): p. 1305-1306.
154. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Research, 2009. **37**: p. W202-W208.
155. Eddy, S.R., *Profile hidden Markov models.* Bioinformatics, 1998. **14**(9): p. 755-63.
156. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* Bmc Bioinformatics, 2010. **11**: p. 119.
157. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-10.
158. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Research, 2004. **14**(6): p. 1188-90.
159. Galtier, N., M. Gouy, and C. Gautier, *SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny.* Computer applications in the biosciences : CABIOS, 1996. **12**(6): p. 543-8.
160. Enz, S., et al., *Analysis of the ferric citrate transport gene promoter of Escherichia coli.* Journal of bacteriology, 2003. **185**(7): p. 2387-91.
161. Lane, W.J. and S.A. Darst, *The structural basis for promoter -35 element recognition by the group IV sigma factors.* PLoS biology, 2006. **4**(9): p. e269.
162. *The filA (rpoF) gene of Pseudomonas aeruginosa encodes an alternative sigma factor required for flagellin synthesis*, 2005. p. 1-12.
163. Potvin, E., F. Sanschagrin, and R.C. Levesque, *Sigma factors in Pseudomonas aeruginosa*, in *FEMS microbiology reviews*2008. p. 38-55.
164. Ellinger, T., et al., *Stalling of Escherichia-Coli Rna-Polymerase in the +6 to +12 Region in-Vivo Is Associated with Tight-Binding to Consensus Promoter Elements.* Journal of Molecular Biology, 1994. **239**(4): p. 455-465.
165. Carpousis, A.J., J.E. Stefano, and J.D. Gralla, *5′ Nucleotide Heterogeneity and Altered Initiation of Transcription at Mutant Lac Promoters.* Journal of Molecular Biology, 1982. **157**(4): p. 619-633.
166. Cunnac, S., M. Lindeberg, and A. Collmer, *Pseudomonas syringae type III secretion system effectors: repertoires in search of functions.* Current Opinion in Microbiology, 2009. **12**(1): p. 53-60.

167. Schechter, L.M., et al., *Functional and Computational Analysis of Amino Acid Patterns Predictive Type III Secretion System Substrates in Pseudomonas syringae.* Plos One, 2012. **7**(4).
168. Baltrus, D.A., et al., *Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 Pseudomonas syringae isolates.* PLoS pathogens, 2011. **7**(7): p. e1002132.
169. Preiter, K., et al., *Novel virulence gene of Pseudomonas syringae pv. tomato strain DC3000.* Journal of bacteriology, 2005. **187**(22): p. 7805-14.
170. Bardoel, B.W., et al., *Pseudomonas Evades Immune Recognition of Flagellin in Both Mammals and Plants.* PLoS pathogens, 2011. **7**(8).
171. Pel, M.J., et al., *Pseudomonas syringae evades host immunity by degrading flagellin monomers with alkaline protease AprA.* Molecular plant-microbe interactions : MPMI, 2014.
172. Létoffé S, D.P., and Wandersman C, *Protease secretion by Erwinia chrysanthemi: the specific secretion functions are analogous to those of Escherichia coli alpha-haemolysin.* EMBO J., 1990. **9**(5): p. 8.
173. Schechter, L.M., et al., *Pseudomonas syringae type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter.* Journal of bacteriology, 2004. **186**(2): p. 543-55.
174. H, N., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of.* - Protein Eng. 1997 Jan;10(1):1-6., (- 0269-2139 (Print)): p. T - ppublish.
175. Wray, G.A., *The evolutionary significance of cis-regulatory mutations.* Nature reviews. Genetics, 2007. **8**(3): p. 206-16.
176. Taguchi, F. and Y. Ichinose, *Role of type IV pili in virulence of Pseudomonas syringae pv. tabaci 6605: correlation of motility, multidrug resistance, and HR-inducing activity on a nonhost plant.* Molecular plant-microbe interactions : MPMI, 2011. **24**(9): p. 1001-11.
177. Damron, F.H., et al., *Analysis of the Pseudomonas aeruginosa Regulon Controlled by the Sensor Kinase KinB and Sigma Factor RpoN.* Journal of bacteriology, 2012. **194**(6): p. 1317-1330.
178. Jones, J., et al., *Integrated bioinformatic and phenotypic analysis of RpoN-dependent traits in the plant growth-promoting bacterium Pseudomonas fluorescens SBW25.* Environmental Microbiology, 2007. **9**(12): p. 3046-64.
179. Tsang, J., et al., *Insertion mutations in Helicobacter pylori flhA reveal strain differences in RpoN-dependent gene expression.* Microbiology, 2013. **159**(Pt 1): p. 58-67.
180. Schulz, T., et al., *FliA expression analysis and influence of the regulatory proteins RpoN, FleQ and FliA on virulence and in vivo fitness in Legionella pneumophila.* Archives of Microbiology, 2012. **194**(12): p. 977-89.
181. Dong, T., R. Yu, and H. Schellhorn, *Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in Escherichia coli.* Molecular Microbiology, 2011. **79**(2): p. 375-86.
182. Zhao, K., M. Liu, and R.R. Burgess, *Promoter and regulon analysis of nitrogen assimilation factor, sigma54, reveal alternative strategy for E. coli MG1655 flagellar biosynthesis.* Nucleic Acids Research, 2010. **38**(4): p. 1273-83.
183. Smith, T.G. and T.R. Hoover, *Deciphering bacterial flagellar gene regulatory networks in the genomic era.* Advances in applied microbiology, 2009. **67**: p. 257-95.

184. Liu, E., et al., *Impact of co-deficiency of RpoN and RpoS on stress tolerance, virulence and gene regulation in Edwardsiella tarda.* Journal of basic microbiology, 2014.
185. Jacobi, S., R. Schade, and K. Heuner, *Characterization of the alternative sigma factor sigma54 and the transcriptional regulator FleQ of Legionella pneumophila, which are both involved in the regulation cascade of flagellar gene expression.* Journal of bacteriology, 2004. **186**(9): p. 2540-7.
186. Kearns, D.B., *A field guide to bacterial swarming motility.* Nature reviews. Microbiology, 2010. **8**(9): p. 634-44.
187. Heurlier, K., et al., *Negative control of quorum sensing by RpoN (sigma54) in Pseudomonas aeruginosa PAO1.* Journal of bacteriology, 2003. **185**(7): p. 2227-35.
188. Thompson, L.S., et al., *The alternative sigma factor RpoN regulates the quorum sensing gene rhlI in Pseudomonas aeruginosa.* Fems Microbiology Letters, 2003. **220**(2): p. 187-95.
189. Sheng, L., et al., *Quorum sensing and alternative sigma factor RpoN regulate type VI secretion system I (T6SSVA1) in fish pathogen Vibrio alginolyticus.* Archives of Microbiology, 2012. **194**(5): p. 379-90.
190. Wang, K., et al., *Characterization of Edwardsiella tarda rpoN: roles in sigma(70) family regulation, growth, stress adaption and virulence toward fish.* Archives of Microbiology, 2012. **194**(6): p. 493-504.
191. Riordan, J.T., et al., *Inactivation of alternative sigma factor 54 (RpoN) leads to increased acid resistance, and alters locus of enterocyte effacement (LEE) expression in Escherichia coli O157 : H7.* Microbiology, 2010. **156**(Pt 3): p. 719-30.
192. Saldias, M.S., et al., *Burkholderia cenocepacia requires the RpoN sigma factor for biofilm formation and intracellular trafficking within macrophages.* Infection and immunity, 2008. **76**(3): p. 1059-67.
193. Stauber, J.L., E. Loginicheva, and L.M. Schechter, *Carbon source and cell density-dependent regulation of type III secretion system gene expression in Pseudomonas syringae pathovar tomato DC3000.* Research in Microbiology, 2012. **163**(8): p. 531-9.
194. Ancona, V., W. Li, and Y. Zhao, *Alternative sigma factor RpoN and its modulation protein YhbH are indispensable for Erwinia amylovora virulence.* Molecular plant pathology, 2014. **15**(1): p. 58-66.
195. Alarcon-Chaidez, F.J., et al., *RpoN (sigma(54)) is required for plasmid-encoded coronatine biosynthesis in Pseudomonas syringae.* Plasmid, 2003. **49**(2): p. 106-17.
196. Buck, M., et al., *The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor.* Journal of bacteriology, 2000. **182**(15): p. 4129-36.
197. Hong, E., M. Doucleff, and D.E. Wemmer, *Structure of the RNA polymerase core-binding domain of sigma(54) reveals a likely conformational fracture point.* Journal of Molecular Biology, 2009. **390**(1): p. 70-82.
198. Juan-Luis Ramos, e.a., *Pseudomonas*. Virulence and Gene Regulation, ed. J.-L. Ramos. Vol. 2. 2004, Granada, Spain: Plenum Publishers.
199. Ishimoto, K.S. and S. Lory, *Formation of pilin in Pseudomonas aeruginosa requires the alternative sigma factor (RpoN) of RNA polymerase.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(6): p. 1954-7.

200. Totten, P.A., J.C. Lara, and S. Lory, *The rpoN gene product of Pseudomonas aeruginosa is required for expression of diverse genes, including the flagellin gene.* Journal of bacteriology, 1990. **172**(1): p. 389-96.
201. Viducic, D., et al., *rpoN gene of Pseudomonas aeruginosa alters its susceptibility to quinolones and carbapenems.* Antimicrobial Agents and Chemotherapy, 2007. **51**(4): p. 1455-62.
202. Hendrickson, E.L., et al., *Differential roles of the Pseudomonas aeruginosa PA14 rpoN gene in pathogenicity in plants, nematodes, insects, and mice.* Journal of bacteriology, 2001. **183**(24): p. 7126-34.
203. Kohler, T., et al., *Involvement of Pseudomonas putida RpoN sigma factor in regulation of various metabolic functions.* Journal of bacteriology, 1989. **171**(8): p. 4326-33.
204. Hendrickson, E.L., et al., *Virulence of the phytopathogen Pseudomonas syringae pv. maculicola is rpoN dependent.* Journal of bacteriology, 2000. **182**(12): p. 3498-507.
205. Pechy-Tarr, M., et al., *RpoN (sigma54) controls production of antifungal compounds and biocontrol activity in Pseudomonas fluorescens CHA0.* Molecular plant-microbe interactions : MPMI, 2005. **18**(3): p. 260-72.
206. Nam, H.S., et al., *The dctA gene of Pseudomonas chlororaphis O6 is under RpoN control and is required for effective root colonization and induction of systemic resistance.* Fems Microbiology Letters, 2006. **256**(1): p. 98-104.
207. Duan, Q., et al., *Flagella and bacterial pathogenicity.* Journal of basic microbiology, 2013. **53**(1): p. 1-8.
208. Apel, D. and M.G. Surette, *Bringing order to a complex molecular machine: the assembly of the bacterial flagella.* Biochimica et biophysica acta, 2008. **1778**(9): p. 1851-8.
209. He, Y., et al., *FliC, a flagellin protein, is essential for the growth and virulence of fish pathogen Edwardsiella tarda.* Plos One, 2012. **7**(9): p. e45070.
210. Abby, S.S. and E.P. Rocha, *The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems.* PLoS genetics, 2012. **8**(9): p. e1002983.
211. Erhardt, M., K. Namba, and K.T. Hughes, *Bacterial nanomachines: the flagellum and type III injectisome.* Cold Spring Harbor perspectives in biology, 2010. **2**(11): p. a000299.
212. Negrea, A., et al., *Salicylidene Acylhydrazides That Affect Type III Protein Secretion in Salmonella enterica Serovar Typhimurium.* Antimicrob Agents Chemother, 2007. **51**(8): p. 10.
213. Vargas, P., et al., *Plant flavonoids target Pseudomonas syringae pv. tomato DC3000 flagella and type III secretion system.* Environmental microbiology reports, 2013. **5**(6): p. 841-850.
214. MacLean, D. and D.J. Studholme, *A Boolean model of the Pseudomonas syringae hrp regulon predicts a tightly regulated system.* Plos One, 2010. **5**(2): p. e9101.
215. Francke, C., et al., *Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior.* BMC genomics, 2011. **12**: p. 385.
216. Chevance, F.F. and K.T. Hughes, *Coordinating assembly of a bacterial macromolecular machine.* Nature reviews. Microbiology, 2008. **6**(6): p. 455-65.

217. Filiatrault, M.J., et al., *CrcZ and CrcX regulate carbon source utilization in Pseudomonas syringae pathovar tomato strain DC3000.* RNA biology, 2013. **10**(2): p. 245-255.
218. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357-9.
219. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology, 1994. **2**: p. 28-36.
220. Haft, D.H., et al., *Genome Properties: a system for the investigation of prokaryotic genetic content for microbiology, genome annotation and comparative genomics.* Bioinformatics, 2005. **21**(3): p. 293-306.
221. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biology, 2010. **11**(10): p. R106.
222. Chambers, A., *Characterization Of The Pseudomonas Syringae Pathovar Tomato Dc3000 Rets Hybrid Two Component Sensor For Induction Of The Type Three Secretion System And Motility.* Master Thesis, 2010.
223. Ortiz-Martín, I., et al., *Negative Regulation of the Hrp Type III Secretion System in Pseudomonas syringae pv. phaseolicola.* MPMI, 2010. **23**(5): p. 10.
224. Bao, Z., et al., *Genomic plasticity enables phenotypic variation of Pseudomonas syringae pv. tomato DC3000.* Plos One, 2014. **9**(2): p. e86628.
225. Lan, L.F., et al., *Genome-wide gene expression analysis of Pseudomonas syringae pv. tomato DC3000 reveals overlapping and distinct pathways regulated by hrpL and hrpRS.* Molecular Plant-Microbe Interactions, 2006. **19**(9): p. 976-987.
226. Grimm, C., W. Aufsatz, and N.J. Panopoulos, *The hrpRS locus of Pseudomonas syringae pv. phaseolicola constitutes a complex regulatory unit.* Molecular Microbiology, 1995. **15**(1): p. 155-65.
227. Heeb, S., C. Blumer, and D. Haas, *Regulatory RNA as mediator in GacA/RsmA-dependent global control of exoproduct.* J Bacteriol., 2002. **184**(4): p. 11.
228. Dasgupta, N., et al., *A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in Pseudomonas aeruginosa.* Molecular Microbiology, 2003. **50**(3): p. 809-824.
229. Jyot, J., N. Dasgupta, and R. Ramphal, *FleQ, the major flagellar gene regulator in Pseudomonas aeruginosa, binds to enhancer sites located either upstream or atypically downstream of the RpoN binding site.* Journal of bacteriology, 2002. **184**(19): p. 5251-60.
230. Starnbach, M.N. and S. Lory, *The fliA (rpoF) gene of Pseudomonas aeruginosa encodes an alternative sigma factor required for flagellin synthesis.* Molecular Microbiology, 1992. **6**(4): p. 459-469.
231. Thilo Ko¨hler, L.K.C., Francisco Barja, Christian Van Delden, And Jean-claude Peche`re, *Swarming of Pseudomonas aeruginosa Is Dependent on Cell-to-Cell Signaling and Requires Flagella and Pili.* J Bacteriol., 2000. **182**(21): p. 7.
232. Murray, T.S. and B.I. Kazmierczak, *FlhF is required for swimming and swarming in Pseudomonas aeruginosa.* Journal of bacteriology, 2006. **188**(19): p. 6995-7004.
233. Daniels, R., J. Vanderleyden, and J. Michiels, *Quorum sensing and swarming migration in bacteria.* FEMS microbiology reviews, 2004. **28**(3): p. 261-89.

234. Lyczak, J.B., C.L. Cannon, and G.B. Pier, *Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist.* Microbes and infection / Institut Pasteur, 2000. **2**(9): p. 1051-60.
235. Hoe, C.H., et al., *Bacterial sRNAs: regulation in stress.* International journal of medical microbiology : IJMM, 2013. **303**(5): p. 217-29.
236. Harris, J.F., et al., *Small RNA-mediated regulation of host-pathogen interactions.* Virulence, 2013. **4**(8): p. 785-795.
237. Caldelari, I., et al., *RNA-mediated regulation in pathogenic bacteria.* Cold Spring Harbor perspectives in medicine, 2013. **3**(9): p. a010298.
238. Sonnleitner, E., A. Romeo, and U. Blasi, *Small regulatory RNAs in Pseudomonas aeruginosa.* RNA biology, 2012. **9**(4): p. 364-71.
239. Burrowes, E., et al., *Influence of the regulatory protein RsmA on cellular functions in Pseudomonas aeruginosa PAO1, as revealed by transcriptome analysis.* Microbiology, 2006. **152**(Pt 2): p. 405-18.
240. Brencic, A. and S. Lory, *Determination of the regulon and identification of novel mRNA targets of Pseudomonas aeruginosa RsmA.* Molecular Microbiology, 2009. **72**(3): p. 612-32.
241. Sonnleitner, E., L. Abdou, and D. Haas, *Small RNA as global regulator of carbon catabolite repression in Pseudomonas aeruginosa.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(51): p. 21866-71.
242. Wiedenheft, B., et al., *RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(25): p. 10092-7.
243. Wenner, N., et al., *NrsZ: a novel, processed, nitrogen-dependent, small non-coding RNA that regulates Pseudomonas aeruginosa PAO1 virulence.* Environmental Microbiology, 2014. **16**(4): p. 1053-68.
244. Gopel, Y., et al., *Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in Enterobacteriaceae.* Nucleic Acids Research, 2011. **39**(4): p. 1294-309.
245. Reichenbach, B., et al., *The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli.* Nucleic Acids Research, 2008. **36**(8): p. 2570-80.
246. Kalamorz, F., et al., *Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli.* Molecular Microbiology, 2007. **65**(6): p. 1518-33.
247. Moreno, R., P. Fonseca, and F. Rojo, *Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in Pseudomonas putida, modulating catabolite repression.* Molecular Microbiology, 2012. **83**(1): p. 24-40.
248. Ermolaeva, M.D., et al., *Prediction of transcription terminators in bacterial genomes.* Journal of Molecular Biology, 2000. **301**(1): p. 27-33.
249. Sharma, A., et al., *Domain movements of the enhancer-dependent sigma factor drive DNA delivery into the RNA polymerase active site: insights from single molecule studies.* (1362-4962 (Electronic)).
250. Kohler, T., J.F. Alvarez, and S. Harayama, *Regulation of the rpoN, ORF102 and ORF154 genes in Pseudomonas putida.* Fems Microbiology Letters, 1994. **115**(2-3): p. 177-84.

251. Iannino, F., R.A. Ugalde, and N. Inon de Iannino, *Characterization of Brucella abortus sigma factor sigma54 (rpoN): genetic complementation of Sinorhizobium meliloti ntrA mutant.* Microbial Pathogenesis, 2008. **45**(5-6): p. 394-402.
252. Clark, S.R.D., I.J. Oresnik, and M.F. Hynes, *RpoN of Rhizobium leguminosarum bv. viciae strain VF39SM plays a central role in FnrN-dependent microaerobic regulation of genes involved in nitrogen fixation.* Molecular and General Genetics, 2001. **264**(5): p. 623-633.
253. Rosenfeld, N., M.B. Elowitz, and U. Alon, *Negative autoregulation speeds the response times of transcription networks.* Journal of Molecular Biology, 2002. **323**(5): p. 785-93.
254. Alejandro Penaloza-Va!zquez, G.M.P., Alan Collmer and Carol L. Bender, *Regulatory interactions between the Hrp type III protein secretion system and coronatine biosynthesis in Pseudomonas syringae pv. tomato DC3000.* Microbiology, 2000. **146**: p. 10.
255. EL, H., et al., *Virulence of the phytopathogen Pseudomonas syringae pv. maculicola is rpoN.* J Bacteriol, 2000. **182**(12): p. 10.
256. Sreedharan, A., et al., *CorR regulates multiple components of virulence in Pseudomonas syringae pv. tomato DC3000.* Molecular plant-microbe interactions : MPMI, 2006. **19**(7): p. 768-79.
257. Bahar, O., T. Goffer, and S. Burdman, *Type IV Pili are required for virulence, twitching motility, and biofilm formation of acidovorax avenae subsp. Citrulli.* Molecular plant-microbe interactions : MPMI, 2009. **22**(8): p. 909-20.
258. Alfano, J.R., et al., *The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(9): p. 4856-61.
259. Jehl, M.A., R. Arnold, and T. Rattei, *Effective--a database of predicted secreted bacterial proteins.* Nucleic Acids Research, 2011. **39**(Database issue): p. D591-5.
260. Gardiner, D.M., K. Kazan, and J.M. Manners, *Nutrient profiling reveals potent inducers of trichothecene biosynthesis in Fusarium graminearum.* Fungal genetics and biology : FG & B, 2009. **46**(8): p. 604-13.
261. Nicaise, V., et al., *Pseudomonas HopU1 modulates plant immune receptor levels by blocking the interaction of their mRNAs with GRP7.* The EMBO journal, 2013. **32**(5): p. 701-12.
262. Jeong, B.R., et al., *Structure function analysis of an ADP-ribosyltransferase type III effector and its RNA-binding target in plant immunity.* The Journal of biological chemistry, 2011. **286**(50): p. 43272-81.
263. Fu, Z.Q., et al., *A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity.* Nature, 2007. **447**(7142): p. 284-8.
264. Ham, J.H., et al., *Multiple activities of the plant pathogen type III effector proteins WtsE and AvrE require WxxxE motifs.* Molecular plant-microbe interactions : MPMI, 2009. **22**(6): p. 703-12.
265. Shine, J. and L. Dalgarno, *Determinant of cistron specificity in bacterial ribosomes.* Nature, 1975. **254**(5495): p. 34-8.

266. Ma, J., A. Campbell, and S. Karlin, *Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures.* Journal of bacteriology, 2002. **184**(20): p. 5733-45.
267. Scharff, L.B., et al., *Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites.* PLoS genetics, 2011. **7**(6): p. e1002155.
268. Kikuchi, G., et al., *Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia.* Proceedings of the Japan Academy. Series B, Physical and biological sciences, 2008. **84**(7): p. 246-63.
269. Boch, J., et al., *Identification of Pseudomonas syringae pv. tomato genes induced during infection of Arabidopsis thaliana.* Molecular Microbiology, 2002. **44**(1): p. 73-88.
270. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic Acids Research, 2014. **42**(Database issue): p. D222-30.
271. Pang, A.S., S. Nathoo, and S.L. Wong, *Cloning and characterization of a pair of novel genes that regulate production of extracellular enzymes in Bacillus subtilis.* Journal of bacteriology, 1991. **173**(1): p. 46-54.
272. Akiyama, M. and H. Nakashima, *Molecular cloning of thi-4, a gene necessary for the biosynthesis of thiamine in Neurospora crassa.* Current genetics, 1996. **30**(1): p. 62-7.
273. Toyama, H., et al., *PqqC/D, which converts a biosynthetic intermediate to pyrroloquinoline quinone.* Biochemical and Biophysical Research Communications, 2002. **299**(2): p. 268-72.
274. Galperin, M.Y. and E.V. Koonin, *A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity.* Protein science : a publication of the Protein Society, 1997. **6**(12): p. 2639-43.